



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

Wouter R. Cox<sup>a,1</sup> , Simone Dobbelaar<sup>b,c</sup> , Martijn Meeter<sup>d</sup> , Merel Kindt<sup>a</sup> , and Vanessa A. van Ast<sup>a,1</sup>

<sup>a</sup>Department of Clinical Psychology, University of Amsterdam, 1001 NK Amsterdam, The Netherlands; <sup>b</sup>Department of Developmental and Educational Psychology, Leiden University, 2311 EZ Leiden, The Netherlands; <sup>c</sup>Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands; and <sup>d</sup>LEARN! Research Institute, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

Edited by Richard G. M. Morris, University of Edinburgh, Edinburgh, Scotland, United Kingdom, and approved October 18, 2021 (received for review January 31, 2021)

**For over a century, stability of spatial context across related episodes has been considered a source of memory interference, impairing memory retrieval. However, contemporary memory integration theory generates a diametrically opposite prediction. Here, we aimed to resolve this discrepancy by manipulating local context similarity across temporally disparate but related episodes and testing the direction and underlying mechanisms of memory change. A series of experiments show that contextual stability produces memory integration and marked reciprocal strengthening. Variable context, conversely, seemed to result in competition such that new memories become enhanced at the expense of original memories. Interestingly, these patterns were virtually inverted in an additional experiment where context was reinstated during recall. These observations 1) identify contextual similarity across original and new memories as an important determinant in the volatility of memory, 2) present a challenge to classic and modern theories on episodic memory change, and 3) indicate that the sensitivity of context-induced memory changes to retrieval conditions may reconcile paradoxical predictions of interference and integration theory.**

episodic memory | spatial context | interference | integration | reconsolidation

**O**ur existing memories can influence how we encode new related experiences, just like new learning can alter how we remember related past events. Related memories that have been encoded at disparate moments are thus not independent of one another. Rather, reciprocal relationships exist between original and more recent episodes that affect the strength of memory recollection. A main challenge of over a century of episodic memory research has been to understand when proactive (from original memories on new ones) and retroactive effects (from new memories on original ones) enhance versus impair memory. In this quest to foretell the fate of memories, many theoretical accounts have pointed to episodic memory's hallmark characteristic: memories are typically encoded and retrieved in relation to a specific spatiotemporal context (1).

In the 20th century, it was soon recognized that when memories overlap (typically in one constituent of a pair of stimuli), retrieval of each memory is impaired (2). However, memory was protected from interference when learning took place across different spatial contexts (3–8; but see ref. 9). Consequently, classic interference theory (2) and related computational models (10) postulate that contextual similarity of events, both proactively and retroactively, leads to episodic memory impairments. This position remains influential in recent theory (11) and is in line with new neuroscientific insights (12).

However, since the return to a previous encoding context is a strong retrieval cue (13–17), the idea that contextual similarity leads to impaired recall diametrically conflicts with memory integration studies showing that new learning is enhanced by retrieval of original associated memories (18–22). Such retrieval can also strengthen existing memories (23–27), as well as

associations across existing and new memories (so-called “inferential memory”) (28–30). Conceivably, earlier studies may not have found enhancements with contextual similarity, because the employed procedures [list-learning in experimental rooms (e.g., ref. 3)] did not explicitly promote associations between learning events and a unique spatial context, and thus failed to induce life-like episodic memories (1) that consist of what-where-when qualities and a recollective experience during recall (31). As a consequence, retrieval during new learning may not have been strong enough to induce memory integration and mnemonic strengthening (19). Therefore, even though contextual similarity has been viewed as a source of interference for decades, it may in fact strengthen, not impair, episodic memory.

Apart from the direction, the underlying mechanisms of context-induced episodic memory change remain unknown as well. Hypothetically, the extent of memory integration during new learning (19), but also the specific conditions during subsequent recall, might be key. For example, as simultaneous retrieval of similar memories has often been suggested to elicit interference (32), memories that have been encoded across the same contexts could show particularly impaired memory recall with contextual reinstatement during retrieval.

Therefore, we performed a series of experiments to elucidate: 1) the direction of context-induced changes of original,

## Significance

**It is not well understood why and how memories of similar events sometimes strengthen each other, but at other times compete, causing either facilitation or impairment of memory recollection. Reasoning that the environment (spatial context) in which events occur is a critical component of memory function, we show that similar memories strengthen each other's retrievability when the events occurred in the same context, whereas memories impair each other's recall when experiences took place across different contexts. Interestingly, these patterns may change or even reverse if contextual cues are present during recall. Our results thus suggest that episodic memory alterations are adjustable rather than permanent, and provide strategies to optimize memory retention and reduce interference in educational and legal settings.**

Author contributions: W.R.C., M.M., and V.A.v.A. designed research; W.R.C. and S.D. performed research; W.R.C., S.D., and V.A.v.A. analyzed data; and W.R.C., S.D., M.M., M.K., and V.A.v.A. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the [PNAS license](#).

<sup>1</sup>To whom correspondence may be addressed. Email: W.R.Cox@uva.nl or V.A.vanAst@uva.nl.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2101509118/-DCSupplemental>.

Published November 24, 2021.

new, and inferential memory [Exp. 1 and Exp. S1 (in *SI Appendix*); 2) the mechanisms (e.g., integration) that underlie these changes (Exp. 2); and 3) whether the effects are different with altered testing conditions (i.e., contextual reinstatement, Exp. 3). We developed a procedure in which spatial context is an integral element of memory episodes, thereby promoting the development of episodic memories (14). Since we were particularly interested in the interactive effects of distinct and established memories, we inserted a delay of 24 h between overlapping learning events (i.e., original memories were consolidated before new learning took place) (33). Specifically, we presented word pairs in unique background images (spatial contexts) (*Materials and Methods*) and instructed participants to vividly visualize scenes involving the meaning of the words in the contexts as a means to create later memories that include what-where-when qualities and a recollective experience during recall (31). Participants visualized related word pairs, such as “Cigar–Book” (AB) and “Cigar–Net” (AC), on separate days (Fig. 1). The word pairs were imagined and tested until participants recalled at least 90% of the B (day 1) and C (day 2) associates correctly. Crucially, on day 2 the pairs were depicted either against the same background contexts as on day 1 (same context condition) or against new contexts (different context condition). For one-third of the original episodes, no new related episode was encoded on day 2, thereby serving as baseline reference for original memories (control). On the third and final day, participants performed cued recall tests for original (AB), new (AC), and inferential memories (BC), without contexts.

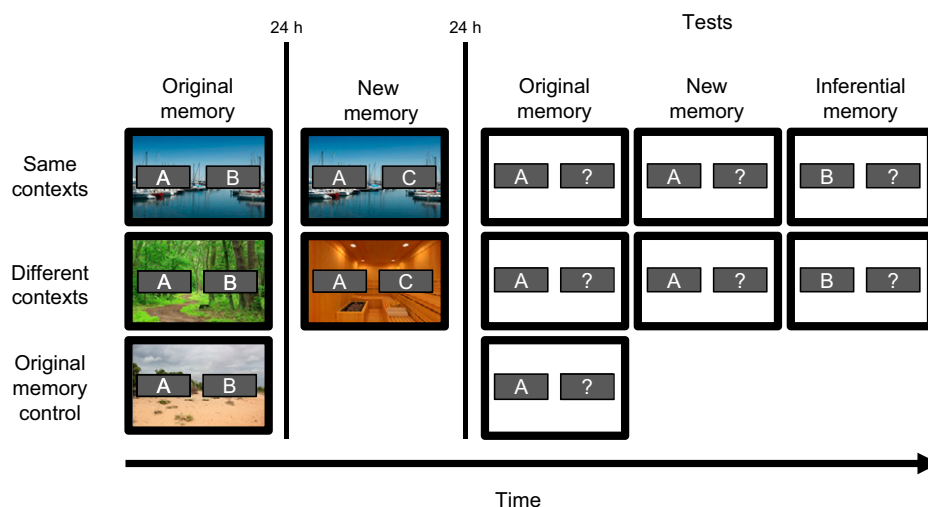
## Results

**Contextual Similarity across Episodes Strengthens Associative Memory, Whereas Contextual Dissimilarity Elicits Interference.** Participants on average needed 1.69 learning rounds ( $SD = 0.66$ ) on day 1 and 1.62 rounds ( $SD = 0.67$ ) on day 2 to correctly recall 90% of the associates. The results of testing on day 3 showed a significant effect of Context on AB accuracy ( $\chi^2_2 = 46.868, P < 0.001$ ) (Fig. 2A). AC learning in the same context markedly enhanced AB memory (same versus control,  $z = 4.922, P < 0.001$ ), while learning in different contexts impaired AB recall (different versus control,  $z = -2.664, P = 0.008$ ). AC

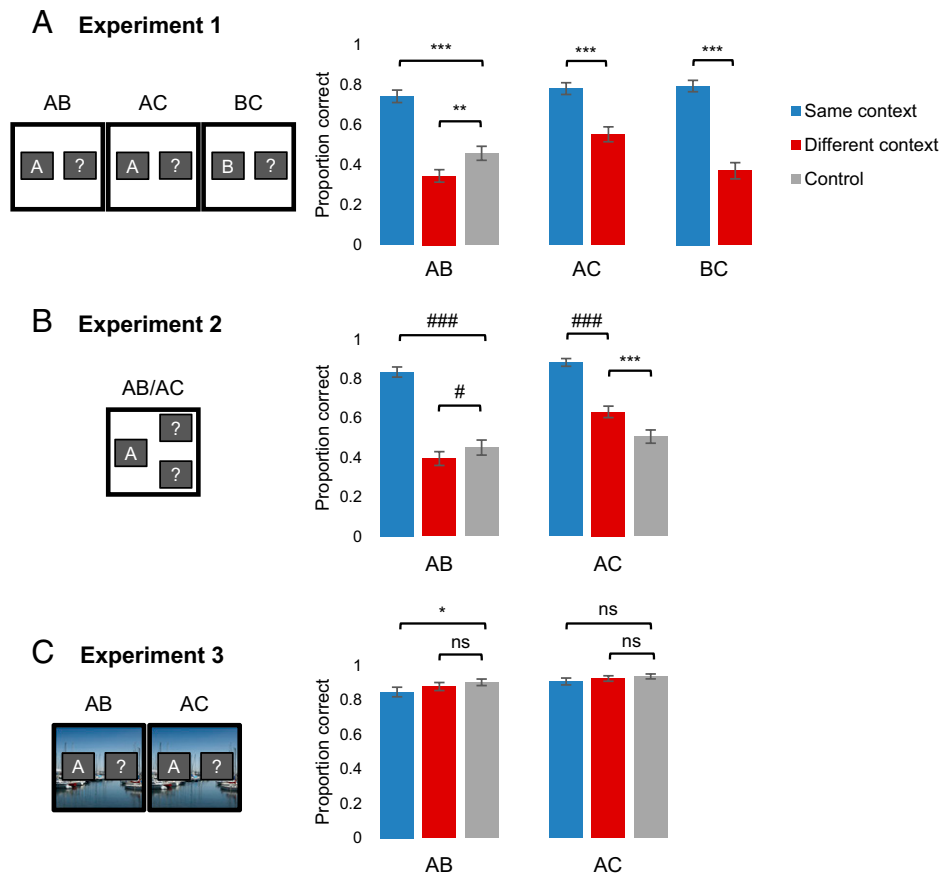
memory and BC inference were relatively enhanced in the same context condition as well (same versus different,  $z = 4.546, P < 0.001$ ;  $z = 5.307, P < 0.001$ , respectively). We also observed an asymmetrical pattern of proactive and retroactive memory misattributions ( $z = -4.236, P < 0.001$ ) (Fig. 3A). That is, C responses during AB testing occurred more often in the different context condition than the same context condition ( $z = -4.737, P < 0.001$ ), while there was no difference between these conditions for B responses during AC testing ( $z = -0.442, P = 0.659$ ). In an additional experiment (Exp. S1; see *SI Appendix*), we reversed the test order, to ascertain that BC results did not carry over from preceding AB and AC tests. The BC findings were replicated (*SI Appendix, Fig. S1*). Note that AB and AC memory could not be validly assessed in Exp. S1 since testing BC first can change performance on these tests (34). Also, given that AB and AC accuracy were markedly different between the same context and different context condition in Exp. 1, and stronger recall of these premise memories very likely facilitates inference, BC performance does not necessarily provide independent evidence for alterations in integration, no matter the testing order.

Together, these findings suggest that contextual similarity across related memories leads to strengthening of episodic memories, whereas contextual dissimilarity induces interference. In addition, asymmetrical patterns of memory misattributions as we observed here have often been attributed to an incorporation of new learning in existing memories through reconsolidation, as opposed to interference from separate memories (35, 36; but see ref. 37). Other changes in episodic memory reconsolidation is thought to be more likely when learning takes place across the same, not different contexts (39). Before assuming reconsolidation as an underlying mechanism in the present study, therefore, additional assessments of the quality of memory change—such as its durability—would be needed (40).

**The Observations Are Best Explained by Effects of Retrieval-Mediated Learning on Memory Recall.** Interestingly, exploratory analyses of Exp. 1 showed that AC accuracy in the same context



**Fig. 1.** Experimental design of Exp. 1. Participants imagined word pairs (AB) in unique context images. One day later, they encoded the same cue word (A) together with a new associate (C). Crucially, a third of the word pairs was presented on the same context image as the day before (same contexts), whereas another third was presented on a new context image (different contexts). For the remaining third of the original episodes, no new related episode was encoded (original memory control). On the final day, participants completed original (AB), new (AC), and inferential (BC) cued recall tests on white backgrounds.



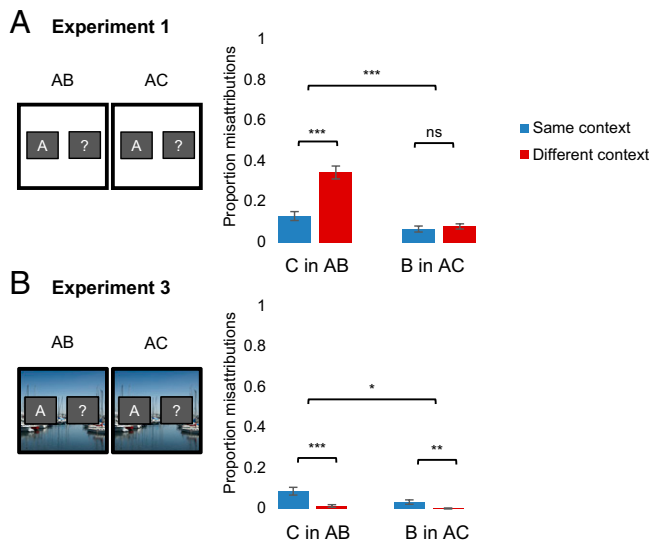
**Fig. 2.** Example trials of testing procedures (Left), and column charts of proportion correct recall per condition (Right). (A) Exp. 1 ( $n = 39$ ). Participants completed separate original (AB), new (AC), and inferential memory (BC) tests. (B) Exp. 2 ( $n = 43$ ). Participants could type in two words in single test trials (42). (C) Exp. 3 ( $n = 41$ ). Participants performed original (AB) and new memory (AC) tests in the respective encoding contexts. Error bars represent SEM. \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ , ### $P < 0.001$  one-sided, # $P < 0.05$  one-sided, ns = not significant.

condition was already higher than the different context condition during new learning. That is, there were no differences in accuracy between conditions during the day 1 testing rounds, but the same context condition showed superior accuracy on day 2 during the first round (*SI Appendix, Fig. S2A*). These findings are in line with memory integration studies showing that enhancement of a new memory that is related to an existing memory, already becomes apparent immediately after encoding of the new memory (retrieval-mediated learning) (18). However, in Exp. 1 it remains unclear whether contextual similarity truly enhanced new learning, since a control condition with novel, nonoverlapping episodes was not included on day 2. Likewise, such a control condition is needed to determine whether new memory is impaired in the different context condition, which would be in line with classic list-learning studies where retroactive and proactive interference typically go hand in hand (41). We, therefore, performed a second experiment in which we added a new memory control condition, controlling for the point in time at which new learning takes place, consisting of 10 completely novel episodes.

In this second experiment, we also aimed to establish whether the effects in Exp. 1 truly reflect changes in memory retrieval. Since recall of original and new episodes was tested separately, accurate memory of day of encoding (source memory) was necessary for correct responding. That is, incorrect responses could have reflected source confusion instead of retrieval inabilities. Therefore, in our second experiment, we adopted a procedure that was designed decades ago to distinguish source confusion from true retrieval decrements (42).

Here, participants could type in both target words (B/C) in response to the cue word (A) in single test trials, thereby excluding source memory influences on the recall scores. They subsequently completed a source memory test of the recalled words (*SI Appendix, Supplementary Methods*).

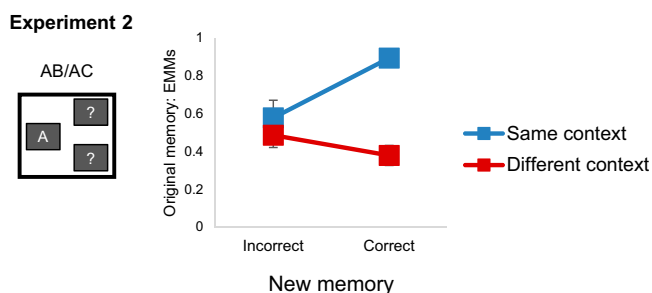
The mean number of learning rounds in Exp. 2 was 1.74 (SD = 0.62) on day 1, and 1.42 (SD = 0.63) on day 2. No differences in accuracy between conditions were observed during day 1 testing, but enhanced recall was found in the same context condition during the first day 2 testing round (*SI Appendix, Fig. S2C*). Even when eliminating influences of source confusion on the recall test on day 3, Exp. 2 replicated the pattern of AB accuracy (Context,  $\chi^2_2 = 62.573$ ,  $P < 0.001$ ; same versus control,  $z = 5.663$ ,  $P < 0.001$ , one-sided; different versus control,  $z = -1.786$ ,  $P = 0.037$ , one-sided) (Fig. 2B). Similarly, AC recall was again bolstered in the same context condition versus the different context condition (Context,  $\chi^2_2 = 62.975$ ,  $P < 0.001$ ,  $z = 5.401$ ,  $P < 0.001$ , one-sided). Crucially, compared to control, we found enhanced AC recall in the same context condition ( $z = 5.667$ ,  $P < 0.001$ ) and, surprisingly, also in the different context condition ( $z = 3.953$ ,  $P < 0.001$ ). These findings confirm that 1) spatial context modulates memory retrieval and not merely source memory (see *SI Appendix, Fig. S3* for source memory accuracy), and 2) that retrieval-mediated learning indeed seems to result in contextual strengthening of new memories. Interestingly, even just slight overlap with original memories (i.e., the A cue) in dissimilar contexts elicited considerable strengthening of new memories.



**Fig. 3.** Example trial of testing procedure (Left), and column chart of proportion of new memory recall during original memory testing (C in AB) and original memory recall during new memory testing (B in AC) per condition (Right). (A) Exp. 1 ( $n = 39$ ). (B) Exp. 3 ( $n = 41$ ). Error bars represent SEM. \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ , ns = not significant.

Note though that the same context condition and the different context condition do not only differ from control in overlap with AB memory, but unavoidably, also in familiarity with the A cue. That is, due to preexposure (during original memory learning) to the A cue in these conditions, A is more familiar. Since familiarity of a retrieval cue can generally boost retrieval of its associate (43), familiarity could have contributed to enhanced new memory recall in the experimental conditions, relative to control. Familiarity does not contribute to changes in the original memories, however, such that these effects can be attributed exclusively to new learning one day later.

**Related Memories Reciprocally Strengthen Each Other or Compete for Persistence, Depending on Contextual Similarity.** Overall, the data suggest that contextual similarity results in memory integration (19) through context-induced retrieval (13–17), enabling strengthened recall of original and new memories. Conversely, contextual dissimilarity induces impaired original memories but strengthened new memories, raising the idea that the memories compete at each other's expense. If this is the case, contextual similarity should lead to a reciprocal strengthening of memories, whereas contextual dissimilarity might induce a trade-off between remembering original and new



**Fig. 4.** Results of a multilevel logistic regression. Original memory accuracy was regressed on correctness of the corresponding new memory trial (correct versus incorrect), and context condition (same context versus different context). Error bars represent SEM. EMMs = estimated marginal means.

memories, on the level of individual episodes. Our previous approach, which involves averages on the subject level, is not suitable to test such interactions between individual memories that can result in changes in specific memories. We therefore performed a multilevel logistic regression with AC accuracy, Context (same context condition and different context condition only), and the unique contexts images in which the AB memories were encoded as predictors and AB accuracy as the outcome. Contextual stability indeed modulated the relationship between AB and AC memory ( $F_{1, 813} = 25.615$ ,  $P < 0.001$ ) (Fig. 4). Planned comparisons confirmed that when learning took place in the same context and AC was recalled, the corresponding AB was more likely to be remembered as well ( $F_{1, 813} = 13.037$ ,  $P < 0.001$ ). Conversely, when learning of AB and AC took place in different contexts, remembering AC resulted in a marginally higher chance of the corresponding AB to not be recalled ( $F_{1, 813} = 3.097$ ,  $P = 0.079$ ).

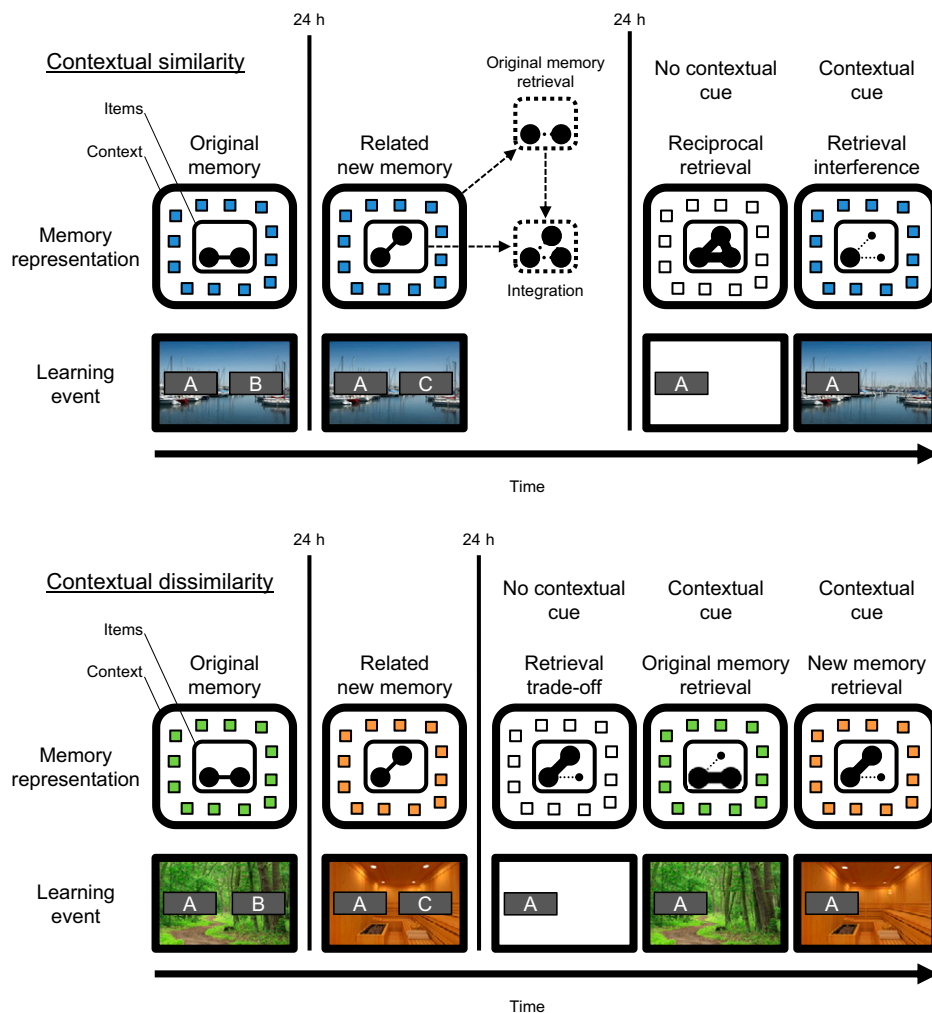
**Competition between Memories Is Markedly Different with Contextual Reinstatement during Recall.** In our third and final experiment, we aimed to assess the context-induced memory changes under different testing conditions. Note that in classic studies showing impaired memory with contextual similarity, memory testing was often performed in the original encoding contexts (3, 4, 6, 7; but see refs. 5 and 8). This leaves open the possibility that contextual similarity of original and new memories was not the sole cause of memory interference, but that stimulating simultaneous memory retrieval by a return to the shared encoding context is a necessary condition for the occurrence of impaired memory accuracy through response competition (32). Conversely, memories that have been encoded in different contexts might be recalled particularly well with contextual reinstatement, also through stimulated retrieval (13–17). In other words, competition between memories might be fundamentally different with contextual reinstatement during recall. Furthermore, probing the memories through contextual reinstatement can shed light on whether the previous findings reflect relatively durable changes in memory, independent of conditions at retrieval (as predicted by reconsolidation theory) (40). Therefore, instead of presenting a white background as before, we now reintroduced the encoding contexts during testing. Note that, unlike in Exp. 2, this necessitates performing separate AB and AC tests, and therefore, source confusion may contribute to the accuracy scores.

The number of learning rounds to recall 90% of the associates was on average 1.73 (SD = 0.67) on day 1 and 1.66 (SD = 0.76) on day 2. The conditions did not differ in accuracy on either day (SI Appendix, Fig. S2D). In sharp contrast with Exp. 1, no difference between the different context and control condition was observed for AB accuracy on day 3 (Context,  $\chi^2_2 = 10.587$ ,  $P = 0.005$ ,  $z = -1.471$ ,  $P = 0.141$ ) (Fig. 2C). Instead, we found impaired accuracy in the same context condition (same versus control,  $z = -2.568$ ,  $P = 0.010$ ). For AC accuracy, no differences between the conditions were found (Context,  $\chi^2_2 = 2.146$ ,  $P = 0.342$ ), although ceiling effects might have occurred. Furthermore, we again observed asymmetrical memory misattributions ( $z = 2.445$ ,  $P = 0.02$ ), but strikingly different from the pattern in Exp. 1. That is, recall of C was prompted during AB testing more frequently in the same context condition than the different context condition ( $z = 3.813$ ,  $P < 0.001$ ) and, to a smaller degree, B during AC testing as well ( $z = 2.970$ ,  $P = 0.003$ ) (Fig. 3B). Thus, context-induced changes in original memory recall, as well as in memory misattributions, seem highly sensitive to retrieval conditions.

## Discussion

Overall, the results reveal that spatial context directs associative memory change (see Fig. 5 for our proposed theoretical





**Fig. 5.** Theoretical model of context-induced changes in episodic memory. Contextual similarity across episodes elicits integration during new learning, leading to bolstered recall of associated memories. However, with contextual reinstatement during recall, retrieval interference (or possibly source confusion) develops instead. Conversely, contextual dissimilarity across episodes induces memory competition, promoting new learning at the expense of original memories. These effects are again overturned by contextual reinstatement, such that accurate memory retrieval occurs instead.

model). Contextual similarity elicits integration of original and new memory episodes, bolstering recall of all involved associations. Contextual dissimilarity, on the other hand, facilitates encoding of original and new episodes into separate memories, resulting in a competitive process that seems to enhance new learning at the expense of original memories. These observations appear to contradict a long history of memory models positing that contextual similarity of memories results in impaired episodic memory while contextual dissimilarity protects against interference (2, 10–12). However, in an additional experiment, where simultaneous retrieval of overlapping memories was stimulated by a return to the original encoding context, contextual similarity in fact resulted in relatively impaired memory accuracy. Likewise, distinct encoding as a result of contextual dissimilarity across episodes no longer induced mnemonic competition when encoding contexts were reinstated during recall. Thus, seemingly paradoxical predictions by recent integration (18–20, 24, 28) and classic forgetting theories (32) may be reconciled by taking into account contextual conditions during retrieval. Therefore, the findings support both a key role for spatial context during new learning, as predicted by memory integration studies, as well as during subsequent recall, in line with interference theory.

Integration theory (18–20, 24, 28), unlike interference (2, 32), does not readily predict that contextual reinstatement during memory testing would relatively impair memories that have been encoded in the same context. Thus, if this observation is replicated in future studies, integration theory may need to accommodate relative worse retrieval of memories learned and retrieved in the same context. Note, however, that it is unclear which specific memories in the same context condition competed and caused this pattern. An interesting possibility is that it was particularly nonintegrated, separate memories that, retrieved through context reinstatement at test (13–17), were subject to response competition and interference (32), while integrated ones were protected from interference (18–20, 24, 28) despite the test in the same context. Furthermore, we cannot rule out that in this experiment source confusion contributed to impaired performance when encoding of original and new memories took place in the same context. It is possible that accuracy was relatively low in the same context condition because the shared context cannot distinguish between the two learning episodes. Consequently, it is harder to remember which of two possible memories belongs to which day of encoding (i.e., source confusion). Thus, memory performance in the same context condition likely does not benefit as much from contextual reinstatement as the different context and control

conditions, where unique associations between a given context and day of encoding exist. For these reasons, it may be worthwhile to further investigate how contextual reinstatement exactly affects integrated memories.

Here, we were particularly interested in the interactive effects of life-like and temporally disparate episodic memories. Therefore, we took care that memories were well-established, by implementing 24 h (including a night of sleep, known to be critical for consolidation) (33) between encoding of overlapping associations. Since learning phases in episodic memory studies on integration (e.g., ref. 18) and interference (e.g., ref. 42) are typically scheduled in quick succession, these studies do not directly allow for conclusions regarding memories that are well-consolidated. For example, the integration elicited by context similarity observed here seems to have promoted “holistic” memory retrieval (e.g., ref. 44), since one successfully retrieved memory made it more likely that a related memory was retrieved as well. However, holistic retrieval typically refers to a memory process that can develop when elements of a single event have been encoded on 1 d (44–46). Likewise, previous studies have suggested that integration seems to develop only when learning events take place on the same day (47). Animal studies indicate that this may be because memory integration critically depends on cell excitability, which decreases already after a few hours (48). Our findings reveal clear evidence for integration of memories that have been encoded across different days, thus suggesting that time is not necessarily a boundary condition for integration to occur, provided that these memories share the same spatial context. Interestingly, this finding is consistent with the observation that cell excitability is reestablished with reexposure to a previously encountered context, which is thought to provide an opportunity for new information to be enduringly associated with a previously formed engram (49). Our study thus opens up avenues to investigate how core episodic memory concepts, such as integration and holistic retrieval, relate to each other and how these may explain mechanisms of memory change well after initial learning in humans.

In contrast with integration or interference theory, the reconsolidation hypothesis does explicitly concern well-consolidated memories. The idea that reactivation of a consolidated memory trace can render it temporarily vulnerable to memory modification (50) has prevalently been suggested to underlie episodic memory change, such as enhancement (51), impairment (38), or incorporation of additional material (35). Several of our observations may seem to align with the reconsolidation hypothesis, but in fact do not fit well with it (40). First, by presenting interfering (AC) information in the same context condition, reconsolidation would predict an impairment in original (AB) memory, whereas memory strengthening was in fact observed. Second, enhancement of new learning by contextual similarity became apparent already during new learning, while reconsolidation-mediated effects should only emerge after a substantial delay (50, 52). Third, reconsolidation theory predicts relatively durable memory modifications, but when memories were contextually probed (in Exp. 3) no interference in the different context condition (as in Exps. 1 and 2) was observed. Likewise, although the asymmetrical memory misattributions when learning and testing take place across the same contexts (Exp. 3) may seem to suggest the occurrence of reconsolidation like in previous studies (e.g., refs. 35, 36, and 39), Exp. 1 shows opposite results. This sensitivity to testing conditions clearly diverges from the idea that new learning would be incorporated in an original memory to form one permanently updated memory upon (context-induced) retrieval (37). Together, these observations suggest that reconsolidation cannot serve as an alternative interpretation of the context-induced retrieval modifications observed here, and furthermore demonstrate the need for stringent behavioral characterization before reconsolidation—or any

other theoretical memory process for that matter—is inferred (40).

The opposing ways in which spatial context during new learning and recall modulate memory retrieval and misattributions is, apart from theoretical considerations, informative at a practical level as well (e.g., in educational and legal settings). The pronounced mnemonic advantage of contextual similarity across episodes when recall takes place outside the original encoding context is, for example, reminiscent of ideas that spatial context provides organizational structure for memories (1, 53), which can be capitalized on by means of the well-known memory palace technique (i.e., method-of-loci) (54). The strength of context-induced memory reactivation is presumably of great importance for the occurrence of this effect, as enhanced memory through integration might occur only, or more robustly, with relatively strong retrieval during new learning (19). However, as the retrieval patterns were markedly different with the presence (Exp. 3) or absence (Exps. 1 and 2) of contextual cues, context during recall seems important as well. Therefore, in any situation where retention of information is essential, the similarity in spatial context across learning episodes, but also subsequent recall, are factors to consider.

Ultimately, the results illustrate spatial context’s ability to adaptively drive memory change. From a functional perspective, contextual stability may indicate that similar original and new events are interrelated and each important to remember, eliciting integration and reciprocal memory strengthening. In contrast, after a contextual change (e.g., moving abroad), it seems functional to remember new events (e.g., first meetings) at the expense of similar older memories for optimal adjustment to changing environments (55). Such retroactive forgetting should, however, not be permanent, as original memories may turn relevant again upon a return to the original environment and should then be retrieved. Thus, the findings exemplify that memories are not fixed. Spatial context signals current or bygone relevance of memories and, thereby, is naturally able to support the flexible nature of episodic memory.

## Materials and Methods

**Exp. 1: The Direction of Context-Induced Memory Change.** In Exp. 1, we tested whether, and in which direction, spatial context may drive changes in temporally disparate associative memories that are linked in content.

**Participants.** Dutch-speaking participants, aged between 18 and 40 y, participated in exchange for course credit or €30. The inclusion criteria were no psychiatric or neurological disorder requiring psychological or pharmacological treatment and no diagnosis of dyslexia. Participants were excluded if they had previously participated in similar experiments or if they were unable to correctly recall at least 90% of the word associates within five learning rounds on day 1 or day 2. The local ethical committee of the University of Amsterdam approved all experimental procedures.

Forty-eight participants consented to take part. Three participants met one of the exclusion criteria (psychiatric disorder,  $n = 2$ ; prior participation in similar research,  $n = 1$ ), five encountered a technical error, and one prematurely ended his participation. Consequently, we analyzed data of 39 participants (27 females) with a mean age of 22 y ( $SD = 2.00$ , range = 18 to 26).

### Stimuli

**Words.** The words for the associative memory task were 90 nouns, obtained from a validated database of Dutch words (56). See *SI Appendix, Supplementary Methods* for a description of the exact selection procedure of individual words, subsequent assemblance of word pairs, and creation of different experimental task versions.

**Contexts.** Context images were in total 40 color pictures of indoor and outdoor spatial environments (e.g., a living room, city landscape), as used in earlier studies (14, 15).

**Procedure.** Participants underwent AB learning (original memory), AC learning (new memory), and memory testing for AB, AC, and BC (inference) on 3 consecutive days (Fig. 1). Upon arrival on day 1, as a cover story, participants were told that the purpose of the experiment was to study practice effects of

an imagination task that would be similar across all 3 experimental days. They were informed that the imagination task involved word pairs and background images. It was stated that on every day the experimental task would consist of two parts: 1) imagining word pairs in a background picture, and 2) testing until imagination of almost all word pairs was successful. We deliberately kept participants unaware that they would perform AB, AC, and BC memory tests on day 3 to preclude active memorizing of the imagined scenes in between sessions. Subsequently, participants gave written informed consent and were screened for the exclusion criteria. Before the start of the experimental task, participants filled in questionnaires. These data were collected for potentially exploratory analyses, but not analyzed.

**AB learning (original memory).** Participants were seated in front of a computer in a dimly lit room, and it was explained that they would be shown word pairs depicted against background images. They were instructed to come up with a vivid scene that involves a visual representation of the words in the environment of the context image while the words and the image were on the screen. This instruction was included to elicit encoding of what-where-when qualities that could result in a recollective experience of the event later on, a typical characteristic of episodic memories (31). To stimulate active imagination, the participants were informed that the imagination phase would be followed by a test. Then, 30 word pairs (AB) in unique context images were presented in total. Each trial (*SI Appendix, Fig. S4A*) started with the background picture being displayed for 2 s, after which the word pair appeared and was shown in the middle of the context image for 6 s. Then the word pair disappeared, and the context was shown alone for an additional 2 s. The next trial started after presentation of a black image for 1 s. The order of trial presentation was semirandom with the restriction that word pairs from one condition (same context, different context, and control) would not be presented more than two times successively.

After imagining all word pairs in their respective backgrounds, participants completed a self-paced memory test (*SI Appendix, Fig. S4B*). It was explained that the word that appeared left on the screen during imagination (A) would be presented together with the context image as a memory cue. Participants were instructed to type in the correct associate (B) and then press the enter key. If they did not remember the correct word, they were asked to type in "xx". Responses would turn green or red, contingent on a correct or incorrect answer, respectively. Then the correct associate (B) was presented together with the cue word (A) and the context image for 4.5 s. In case of a wrong answer, participants were instructed to again imagine the word pairs in the background image, like in the initial imagination phase. The testing procedure was repeated for the complete list of word pairs until 90% of the associates (B) were correctly recalled, up to five test rounds maximum. Reaction times were recorded during all recall tests. The order of trial presentation was again semirandomized, with the same restrictions as during learning.

**AC learning (new memory).** Twenty-four hours later, participants imagined 20 new word combinations in context images. The word pairs contained the same cue word (A) as the day before, but a new associate (C). Crucially, 10 of the AC pairs were presented in the same context as the corresponding AB word pair (same context condition), whereas the other 10 pairs were displayed on a new background image (different context condition). For the 10 remaining word pairs from the AB learning phase, no new related episode was encoded, thereby serving as baseline reference for the point in time at which original memories were encoded (original memory control condition). All other AC learning instructions and testing procedures were identical to AB learning.

**AB, AC, and BC memory tests.** On the third and final day, participants completed self-paced AB (original memory), AC (new memory), and BC (inference) tests, in that order. For the AB test, it was explained to participants that the cue words (A) would appear on a white background, and they had to type in the corresponding day 1 associate (B). After completion of all AB test trials, they were told that the cue words (A) would appear again, but now the day 2 associates (C) had to be filled in. Subsequently, to assess inference, participants were made aware that they had learned two associates (B and C) in combination with the same cue (A) on the previous days. It was explained that they would be presented with the B associate next and had to recall the corresponding C associate.

For all tests, the trials were again presented in a quasi-random order, such that cue words from one condition would not be presented more than two times consecutively. No feedback was provided during memory testing.

Upon completion of the BC test, participants were interviewed regarding encoding characteristics, such as learning strategies, general vividness of their imagined scenes, and use of the background picture during imagination. Finally, participants were debriefed and rewarded for participation.

**Data Analysis.** All individual trials were inspected independently by two authors (W.R.C. and S.D.). Misspellings, singularity versus plurality mistakes,

and diminutives of the target word were treated as correct responses. The raters reached consensus on items at which their scores differed (proportion agreement = 0.94 across experiments). The participants' proportion of correct responses during AB (original memory), AC (new memory), and BC (inference) testing were then calculated per condition (i.e., same context, different context, control). In addition, the proportion of correct responses during day 1 and day 2 learning were calculated for all testing rounds. Finally, we calculated the proportion of new memory misattributions (C) during the original memory test (AB), and original memory misattributions (B) during the new memory test (AC), for the same context and different context conditions.

Statistical analyses were carried out using the Statistical Package for the Social Sciences v25.0 (IBM Corp.). As the data were nonnormally distributed, AB memory accuracy scores were analyzed through a Friedman's ANOVA with Context (same context, different context, control) as within-subjects factor. Wilcoxon signed-rank tests were used to perform planned comparisons of specific conditions. Because of the relatively large sample size, a z-statistic was used rather than *W*. Also, AC and BC memory accuracies were analyzed using Wilcoxon signed-rank tests. Accuracy scores of day 1 and day 2 testing were analyzed for each testing round separately, in the same way as day 3 accuracies.

As the proportion of B responses during AC testing and C responses during AB testing (memory misattributions) were nonnormally distributed as well, we calculated difference scores of misattributions in the different context condition relative to the same context condition per memory type (AB, AC). These normalized scores were compared using a Wilcoxon signed-rank test to test for a Context  $\times$  Memory interaction. Wilcoxon signed-rank tests were also used to analyze differences between the same context and different context conditions in misattributions during AB and AC memory testing separately. We performed two-sided tests.  $\alpha$  was set at 0.05.

**Exp. 2: The Nature of Context-Induced Memory Change.** To 1) uncover whether new memories that have been encoded in same contexts as original memories are enhanced, whereas contextual dissimilarity leads to impairment of new memories, and 2) establish whether the effects in Exp. 1 truly reflect changes in memory retrieval, we changed two aspects of the procedure relative to Exp. 1. We 1) included a control condition for new memories, and 2) had participants type in both target words (B/C) in response to the cue word (A) in single test trials, and subsequently perform a source memory test of the recalled words. This is a classic testing procedure to assess true retrieval decrements independent of source confusion (42).

**Participants.** Forty-eight participants consented to take part. Three participants met one of the exclusion criteria (prior participation in similar research,  $n = 2$ ; unable to recall 90% of word associates within five learning rounds on day 2,  $n = 1$ ). Additionally, two participants encountered a technical error, leaving 43 participants (34 females) with a mean age of 21 y ( $SD = 2.13$ , range = 18 to 26) for analyses.

#### Stimuli

**Words.** To create a control condition for the new memories, we selected 30 additional words from the same database as before (56) (*SI Appendix, Supplementary Methods*).

**Contexts.** Ten context images were added to complete the new memory control condition.

#### Procedure

**AB learning (original memory).** All AB learning instructions and testing procedures were the same as in Exp. 1.

**AC learning (new memory).** Participants imagined in total 30 word combinations in context images. As in Exp. 1, 20 word pairs contained the same cue word (A) as on day 1, with a new associate (C). Again, half of these pairs were presented in the same context as the corresponding AB word pair (same context condition), whereas the other half was displayed on a new background image (different context condition). But, crucially, for the added new memory control condition (i.e., controlling for the point in time at which new learning takes place), participants imagined 10 additional word combinations. These pairs consisted of both a new cue and a new associate, and were displayed with a new context. All learning instructions and testing procedures were the same as during AB learning.

**AB and AC memory tests.** On the final day, participants completed self-paced cued recall of the B and C associates on white backgrounds within single test trials (so-called "modified modified free recall") (42). It was explained to participants that at each trial, they would be shown the cue word (A) with two empty text boxes next to it (Fig. 2B). To determine retrieval of the target words, they were instructed to type in the B and C associates as they came to mind. Importantly, this testing procedure guaranteed that responses were



independent of source memory, as it did not require remembrance of the specific day of encoding. It was explained to them that they could continue to the next trial only when they had filled in both text boxes. Participants were also informed that for some trials they only had learned one associate (i.e., for the control conditions). For these cases, and trials at which they were unable to recall the target word, they were instructed to fill in "xx" in one of the text-boxes. Other testing procedures were the same as in Exp. 1. No BC (inference) tests were performed. After participants had completed memory recall, they performed a self-paced source memory test, during which they were presented with the A cue and the words they had just recalled. To assess source memory accuracy, they were asked to tick one of three boxes (day 1, day 2, do not know) that were displayed next to every recalled word (see *SI Appendix, Supplementary Methods* for more details).

**Data Analysis.** To obtain AB and AC memory accuracy scores that reflect only memory retrieval and are not taxed by source memory, we calculated the proportion of correct B and C associates that were retrieved in response to the corresponding cue word (A) during the recall test or came to mind during source memory testing, as our main outcome variable. We also ran an analysis (which yielded the same conclusions) to control for a potentially different number of response opportunities between conditions, and two analyses to assess source memory accuracy (see *SI Appendix, Supplementary Methods* for more details). The accuracy scores were analyzed using a Friedman's ANOVA with Context as a within-subjects factor. Planned comparisons were performed with Wilcoxon signed-rank tests. Since we predicted the direction of the effects to replicate those observed in Exp. 1, one-sided tests were performed. As no new memory control condition was included in Exp. 1, two-sided tests were used when a comparison was made with this specific new condition.

Furthermore, to assess whether memories are truly reciprocally connected, such that in the same context condition memories are more likely to strengthen each other, whereas in the different context condition memories persist at each other's expense, it is essential to additionally test how memories related to each other specifically at the memory—instead of the subject—level. We therefore performed a multilevel logistic regression of individual memory episodes. On the item level (first level), we entered AC (new memory) accuracy (correct, incorrect) as a predictor and AB (original memory) accuracy (correct, incorrect) as the outcome. It is possible that certain specific context images provide a particularly potent scaffold to build memories, resulting in strong recall of AB memory. Such effects of specific context images could mask interference or cause enhanced performance regardless of contextual similarity. Hence, to control for possible effects of individual background images that could confound the result, we also added the unique context images that AB were encoded in as a fixed factor (individual image identifiers). At the participant level (second level), we entered context condition (same context or different context) to test for modulation of the relation

between AC and AB recall. AC and AB accuracy were entered as ordinal variables, and context image and context condition as nominal variables. A diagonal covariance matrix was used. As inclusion of a random slope (AC relationship with AB performance) did not explain any additional variance, the model was random intercept (participant) only (57). Otherwise, we entered all theoretical predictors at once and did not further engage in any model selection procedures.

**Exp. 3: Context-Induced Memory Change with Contextual Reinstatement during Recall.** Performing retrieval in the encoding context might lead to relatively impaired recall of episodes that have been encoded in same contexts due to simultaneous retrieval of similar memories (32). Conversely, memory for episodes containing unique contexts might be facilitated by contextual reinstatement (13–17). To assess whether competition between memories is different with contextual reinstatement during recall, we therefore presented the original encoding contexts, instead of a white background, during the memory tests on day 3. This approach also enabled to test whether the previously observed changes in recall are durable (as would be predicted by reconsolidation theory) (40).

**Participants.** Forty-five participants consented to take part. Two participants met one of the exclusion criteria (psychiatric disorder) and two encountered a technical error. Consequently, we analyzed data of 41 participants (34 females) with a mean age of 21 y (SD = 4.01, range = 18 to 40).

**Stimuli.** The same stimuli were used as in Exp. 2.

**Procedure.** All AB and AC learning instructions and testing procedures were the same as in Exp. 2.

**AB and AC memory tests.** On the third day, participants completed separate original (AB) and new (AC) memory tests as in Exp. 1. However, now the cue words (A) were displayed in their original encoding contexts, instead of on a white background (Fig. 2C). No BC (inference) tests were performed.

**Data Analysis.** The data analysis strategies were the same as in Exp. 1.

**Data Availability.** The raw and processed data have been deposited in the Open Science Framework (OSF) and are available at <https://osf.io/qj2sa/>.

**ACKNOWLEDGMENTS.** The authors thank Heleen Bes, Lara Bridge, Nikki Evers, and Allerd Reitsma for help with data collection; Raoul Grasman for statistical advice on the multilevel logistic analysis; Bert Molenkamp for technical assistance; and the anonymous reviewers for their thoughtful comments. This work was supported by Research Talent Grant 406-16-557 (to W.R.C. and M.K.) and Veni Grant 451-16-021 (to V.A.v.A.) from the Netherlands Organization for Scientific Research. M.K. is supported by a European Research Council Advanced Grant 743263.

1. E. Tulving, "Episodic and semantic memory" in *Organization of Memory*, E. Tulving, W. Donaldson, Eds. (Academic Press, 1972), pp. 382–402.
2. J. A. McGeoch, Forgetting and the law of disuse. *Psychol. Rev.* **39**, 352–370 (1932).
3. I. M. Bilodeau, H. Schlosberg, Similarity in stimulating conditions as a variable in retroactive inhibition. *J. Exp. Psychol.* **41**, 199–204 (1951).
4. K. Dallett, S. G. Wilcox, Contextual stimuli and proactive inhibition. *J. Exp. Psychol.* **78**, 475–480 (1968).
5. J. Greenspoon, R. Ranyard, Stimulus conditions and retroactive inhibition. *J. Exp. Psychol.* **53**, 55–59 (1957).
6. B. Z. Strand, Change of context and retroactive inhibition. *J. Verb. Learn. Verb. Be.* **9**, 202–206 (1970).
7. L. Jensen, J. Dibble, D. C. Anderson, Effects of a contextual change upon retroactive inhibition. *Psychol. Rep.* **29**, 39–46 (1971).
8. S. M. Smith, Enhancement of recall using multiple environmental contexts during learning. *Mem. Cognit.* **10**, 405–412 (1982).
9. J. W. Nagge, An experimental test of the theory of associative interference. *J. Exp. Psychol.* **18**, 663–682 (1935).
10. G. Mensink, J. G. W. Raaijmakers, A model for interference and forgetting. *Psychol. Rev.* **95**, 434–455 (1988).
11. A. P. Yonelinas, C. Ranganath, A. D. Ekstrom, B. J. Wiltgen, A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nat. Rev. Neurosci.* **20**, 364–375 (2019).
12. D. M. Smith, D. A. Bulkin, The form and function of hippocampal context representations. *Neurosci. Biobehav. Rev.* **40**, 52–61 (2014).
13. S. M. Smith, E. Vela, Environmental context-dependent memory: A review and meta-analysis. *Psychon. Bull. Rev.* **8**, 203–220 (2001).
14. 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).
15. 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).
16. D. R. Godden, A. D. Baddeley, Context-dependent memory in two natural environments: On land and underwater. *Br. J. Psychol.* **66**, 325–331 (1975).
17. S. M. Smith, Remembering in and out of context. *J. Exp. Psychol. Hum. Learn.* **5**, 460–471 (1979).
18. M. L. Schlichting, A. R. Preston, Memory reactivation during rest supports upcoming learning of related content. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15845–15850 (2014).
19. A. J. H. Chanales, N. M. Dudukovic, F. R. Richter, B. A. Kuhl, Interference between overlapping memories is predicted by neural states during learning. *Nat. Commun.* **10**, 5363 (2019).
20. C. N. Wahlheim, Testing can counteract proactive interference by integrating competing information. *Mem. Cognit.* **43**, 27–38 (2015).
21. L. L. Jacoby, C. N. Wahlheim, On the importance of looking back: The role of recursive reminders in recency judgments and cued recall. *Mem. Cognit.* **41**, 625–637 (2013).
22. K. K. Szpunar, K. B. McDermott, H. L. Roediger III, Testing during study insulates against the buildup of proactive interference. *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 1392–1399 (2008).
23. J. D. Karpicke, H. L. Roediger III, The critical importance of retrieval for learning. *Science* **319**, 966–968 (2008).
24. B. A. Kuhl, A. T. Shah, S. DuBrow, A. D. Wagner, Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* **13**, 501–506 (2010).
25. D. Oudiette, J. W. Antony, J. D. Creery, K. A. Paller, The role of memory reactivation during wakefulness and sleep in determining which memories endure. *J. Neurosci.* **33**, 6672–6678 (2013).
26. J. D. Koen, M. D. Rugg, Memory reactivation predicts resistance to retroactive interference: Evidence from multivariate classification and pattern similarity analyses. *J. Neurosci.* **36**, 4389–4399 (2016).
27. R. Potts, D. R. Shanks, Can testing immunize memories against interference? *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 1780–1785 (2012).



28. D. Zeithamova, A. L. Dominick, A. R. Preston, Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron* **75**, 168–179 (2012).
29. M. T. R. van Kesteren, P. Rignanesi, P. G. Gianferrara, L. Krabbendam, M. Meeter, Congruency and reactivation aid memory integration through reinstatement of prior knowledge. *Sci. Rep.* **10**, 4776 (2020).
30. M. T. R. van Kesteren, L. Krabbendam, M. Meeter, Integrating educational knowledge: Reactivation of prior knowledge during educational learning enhances memory integration. *NPJ Sci. Learn.* **3**, 11 (2018).
31. E. Tulving, Episodic memory: From mind to brain. *Annu. Rev. Psychol.* **53**, 1–25 (2002).
32. M. C. Anderson, Rethinking interference theory: Executive control and the mechanisms of forgetting. *J. Mem. Lang.* **49**, 415–445 (2003).
33. R. Stickgold, Sleep-dependent memory consolidation. *Nature* **437**, 1272–1278 (2005).
34. A. C. Carpenter, D. L. Schacter, Flexible retrieval: When true inferences produce false memories. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 335–349 (2017).
35. A. Hupbach, R. Gomez, O. Hardt, L. Nadel, Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learn. Mem.* **14**, 47–53 (2007).
36. A. Hupbach, R. Gomez, L. Nadel, Episodic memory reconsolidation: Updating or source confusion? *Memory* **17**, 502–510 (2009).
37. A. Klingmüller, J. B. Caplan, T. Sommer, Intrusions in episodic memory: Reconsolidation or interference? *Learn. Mem.* **24**, 216–224 (2017).
38. C. Forcato *et al.*, Reconsolidation of declarative memory in humans. *Learn. Mem.* **14**, 295–303 (2007).
39. A. Hupbach, O. Hardt, R. Gomez, L. Nadel, The dynamics of memory: Context-dependent updating. *Learn. Mem.* **15**, 574–579 (2008).
40. J. W. B. Elsey, V. A. Van Ast, M. Kindt, Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychol. Bull.* **144**, 797–848 (2018).
41. R. J. Seidel, The concurrent effects of proactive and retroactive inhibition. *J. Exp. Psychol.* **57**, 397–402 (1959).
42. J. M. Barnes, B. J. Underwood, Fate of first-list associations in transfer theory. *J. Exp. Psychol.* **58**, 97–105 (1959).
43. M. Hanczakowski, T. Pasek, K. Zawadzka, G. Mazzoni, Cue familiarity and “don’t know” responding in episodic memory tasks. *J. Mem. Lang.* **69**, 368–383 (2013).
44. A. J. Horner, N. Burgess, The associative structure of memory for multi-element events. *J. Exp. Psychol. Gen.* **142**, 1370–1383 (2013).
45. A. J. Horner, N. Burgess, Pattern completion in multielement event engrams. *Curr. Biol.* **24**, 988–992 (2014).
46. A. J. Horner, J. A. Bisby, D. Bush, W.-J. Lin, N. Burgess, Evidence for holistic episodic recollection via hippocampal pattern completion. *Nat. Commun.* **6**, 7462 (2015).
47. D. Zeithamova, A. R. Preston, Temporal proximity promotes integration of overlapping events. *J. Cogn. Neurosci.* **29**, 1311–1323 (2017).
48. S. A. Josselyn, P. W. Frankland, Memory allocation: Mechanisms and function. *Annu. Rev. Neurosci.* **41**, 389–413 (2018).
49. M. Pignatelli *et al.*, Engram cell excitability state determines the efficacy of memory retrieval. *Neuron* **101**, 274–284.e5 (2019).
50. K. Nader, G. E. Schafe, J. E. Le Doux, Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* **406**, 722–726 (2000).
51. C. Forcato, M. L. C. Rodriguez, M. E. Pedreira, Repeated labilization-reconsolidation processes strengthen declarative memory in humans. *PLoS One* **6**, e23305 (2011).
52. M. Kindt, M. Soeter, Pharmacologically induced amnesia for learned fear is time and sleep dependent. *Nat. Commun.* **9**, 1316 (2018).
53. H. Eichenbaum, Memory: Organization and control. *Annu. Rev. Psychol.* **68**, 19–45 (2017).
54. I. C. Wagner *et al.*, Durable memories and efficient neural coding through mnemonic training using the method of loci. *Sci. Adv.* **7**, eabc7606 (2021).
55. H. L. Roediger, Y. Weinstein, P. K. Agarwal, “Forgetting: Preliminary considerations” in *Forgetting*, S. Della Sala, Ed. (Psychology Press, 2010), pp. 1–22.
56. 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).
57. J. W. R. Twisk, *Applied Multilevel Analysis: A Practical Guide* (Cambridge University Press, 2006).