The time window of reconsolidation: A replication

Colleen M. Parks1 · Kevin D. Mohawk1 · Laura L. S. Werner1 · Christopher Kiley1

Accepted: 8 April 2022 / Published online: 15 April 2022
© The Psychonomic Society, Inc. 2022

Abstract

Reconsolidation is a process by which a consolidated memory that has been destabilized by reactivation is updated, strengthened, or weakened by the restabilization of the trace. A critical assumption of the reconsolidation theory is that reconsolidation is a time-dependent process. Hupbach, Gomez, Hardt, and Nadel (2007, Learning & Memory, 14, 47–53) conducted a set of experiments demonstrating that memory updating is only found when the reconsolidation process has time to complete. This finding strengthens reconsolidation theory and poses a challenge to other accounts of memory updating (e.g., context and interference accounts). Because this finding is so critical to the reconsolidation theory, we attempted to directly replicate these experiments, which showed memory updating in a 3-day paradigm (when reconsolidation has time to complete), but not in a 2-day paradigm (when reconsolidation does not have time to complete). We replicated these results, thereby bolstering the reconsolidation theory of memory updating.

Keywords    Reconsolidation · Memory updating · Episodic memory

According to reconsolidation theory, retrieval of an existing memory trace can destabilize it, which in turn opens a time-dependent window during which the memory becomes malleable. The memory can then be altered, strengthened, or weakened (Inda et al., 2011; see Elsey et al., 2018, for a review). In one of the first studies to examine reconsolidation in human episodic memory, Hupbach et al. (2007) found that consolidated memories were updated with new information when participants were reminded of the original learning episode. Specifically, participants memorized a set of objects (balloon, sock, etc.) to an 85% criterion in one session. Two days later, one group of participants was reminded of the Day 1 procedure by returning to the same spatial context as Day 1 and being asked to recall the general procedure of Day 1. A no-reminder group was taken to a different space on Day 2 and was not asked to recall the procedure of Day 1. Participants in the reminder and no-reminder groups learned another set of objects to criterion on Day 2. On Day 3 (another 2 days later), both groups were given a free recall test of Day 1 items; a control group that did not participate in Day 2 was also tested. Results showed that participants who received a reminder of Day 1 before learning List 2 made significantly more List-2 intrusions than did the no-reminder and control groups. This memory updating was attributed to reconsolidation, which is argued to be a unique neurobiological process.

There are challenges to the reconsolidation theory, though (Capelo et al., 2018; Gisquet-Verrier et al., 2015; Gisquet-Verrier & Riccio, 2018; Kiley & Parks, 2022; Klingmüller et al., 2017; Sederberg et al., 2011), most of which argue that the memory updating found in Hupbach et al.’s (2007) paradigm is attributable to context-related retroactive interference or encoding specificity and not a distinct neurobiological process. However, these alternatives struggle to explain the time-dependent nature of memory updating (but see Sederberg et al., 2011). That is, because reconsolidation is a process that is supposed to take time, much like consolidation, reconsolidation theory holds that the memory-updating pattern should only be found after a delay between Day 2 learning and the final test on Day 3 (e.g., 24 or 48 hours). However, a context hypothesis does not specify a time window and has difficulty accounting for its necessity. For example, it has been argued that intrusions due to context-based retroactive interference should occur regardless of when the test is administered (Hupbach et al., 2007). To pit these two hypotheses against each other, Hupbach et al. (2007) conducted a second experiment that was identical to
the first, with the exception that the final test (over Day-1 objects) was given immediately after learning Day-2 objects. Thus, there was no time for the reconsolidation process to unfold. They found no difference in intrusions between the reminder and no-reminder conditions, and this null effect was taken as strong evidence in favor of time dependency, and strong evidence in favor of reconsolidation itself (see Hupbach et al., 2011, for a similar result in children). Thus, this finding is critically important to the reconsolidation hypothesis. Had the memory-updating pattern been found in this experiment, it would have been strong evidence in favor of a context and interference account.

Because time dependency is so crucial to reconsolidation theory, and because it has not been thoroughly tested by independent labs, we directly replicated Hupbach et al.’s (2007) Experiments 1 and 2. Experiment 1 was the standard 3-day paradigm and Experiment 2 was a 2-day paradigm, as described above. If the original results replicate, there should be more intrusions from Day 2 on the test of Day 1 in the reminder condition than in the no-reminder condition for the 3-day experiment, but this effect should be eliminated in the 2-day experiment.

Method

Participants

Effect sizes were not reported in Hupbach et al. (2007), and therefore the power analysis was based on the effect sizes reported in a meta-analysis of reconsolidation in episodic memory (Scully et al., 2017). The meta-analysis of eight effect sizes of List-2 intrusions on List-1 memory produced a mean effect size of Cohen’s $d = 1.03$. With that effect size and power set at $.95$, $G^*$Power 3.0 indicated that we needed 22 participants per condition (a total of 132 for both experiments). Participants were UNLV students (female $N = 95$; mean age = 22.11, $SD = 6.55$) who participated for course credit or for $10$ an hour.

Materials

Stimuli consisted of 40 everyday objects (see supplementary material for the complete list), split into two lists for the 2 days, with 20 randomly assigned to the two lists. Some of the objects differed from those used by Hupbach et al. (2007) so that we could use objects that were easily named with one word.

Procedure

Experiment 1 The first day was the same across all groups in both experiments. Participants were presented with a set of objects that they named and tried to remember for a later memory test. Objects were taken out of a bag one at a time, and after the participant named the object, it was placed in a blue basket. A free recall test was given immediately after the presentation of the objects. Participants went through this study–test procedure until they recalled 17 of the 20 objects (85%) or until they completed four study–test blocks. At the end of Day 1, participants in the reminder group were told to come back to the lab 48 hours later. Participants in the no-reminder group were told that they would meet a different experimenter outside the library (different space than Day 1) 48 hours later. Due to COVID-19 disruptions, some participants in this condition could not be tested at the university library and thus were tested in a different building, still distinct from the lab setting. Participants in the no-contact control group were told to return to the lab for Day 3 (96 hours later).

On Day 2, participants in the reminder group returned to the lab, participants in the no-reminder group went to the university library, and participants in the interference control condition skipped Day 2 entirely. The reminder group was tested in the same room and by the same experimenter as Day 1. They were shown the blue basket and asked to describe the procedure on Day 1 (“Do you remember this blue basket and what we did with it?”) but were stopped if they began to recall the objects. The no-reminder group was tested in a study space in the library (or another building) with a new experimenter and was not asked about Day 1. In both groups, the procedure for studying the new series of objects differed from Day 1 so as not to serve as a reminder. The to-be-remembered objects were laid out all at once; participants were asked to name each object and were then given 30 s to study them. The objects were then covered, and a free recall test began; participants went through this study–test procedure until they reached criterion or until they completed four study–test blocks.

On Day 3, all participants returned to the lab to take a free recall test over Day 1 objects. At the point that the participant indicated that they could not remember any more items, the experimenter engaged the participant in a conversation about an unrelated topic for 30 seconds. The experimenter repeated the recall test until the participant could not remember any more items. Participants took the test four times in total.

Experiment 2 Day 1 followed the same procedure as that described for Experiment 1. On Day 2, the reminder group returned to the lab to learn another list and were then tested; the no-reminder group met a different experimenter at the library (or a different building), and the control group returned to the lab. The learning procedure was the same as that used in Experiment 1 (for the reminder and no-reminder groups); however, immediately after learning the objects,
participants were given the test over the Day 1 objects. Those in the interference control group were given a test only over Day 1 objects (i.e., no Day 2 learning). Those in the reminder group took a 5-minute break between studying List 2 objects and the test over List 1 objects (in order to equate them to the no-reminder group who had a break to walk back to the lab from the different setting). Those in the no-reminder group went from the library (or different building) to the lab for the Day 1 test. Those in the control group were immediately given the Day-1 test (no Day-2 learning). Experiment 2 was complete on Day 2. On the last day of the testing in both experiments, participants were given an exit interview (see Supplementary Materials). Exploratory analysis of responses to these questions revealed no trends that would be problematic for interpreting the results of the study.

**Deviations from the original study**

The only differences between this study and Hupbach et al. (2007) were some of the specific objects used, the inclusion of more participants, and the use of an exit interview.

**Results**

For both experiments, we conducted analyses of variance (ANOVA) and follow-up *t* tests to analyze the results, which included three dependent variables (acquisition on Days 1 and 2, accurate recall of List 1, and intrusions from List 2 into List 1 memory). We did this both with significance testing to replicate the approach that Hupbach et al. (2007) used and with Bayesian analyses. We conducted one test that Hupbach et al. did not use, a comparison of the results across the experiments.

Hupbach et al. (2007) did not report the exclusion of data. We excluded and replaced two subjects due to their self-report of a memory impairment. In a supplementary analysis, outliers, defined as two or more standard deviations away from the mean, were removed and the tests were conducted again. We footnote those cases where outliers affected whether a result was significant or not. Below, we report both significance tests, with alpha set at .05, and Bayes factors (BF) for the alternative hypothesis (BF10). For effect sizes, we report Cohen’s *d* for *t* tests and partial eta squared (ηp2) for ANOVAs. We use Greenhouse–Geisser corrections when the sphericity assumption was violated.

**Acquisition of Lists on Days 1 and 2** For Experiment 1 (3-day conditions), the number of blocks taken to reach a criterion of 85% accuracy was compared across groups to determine whether there were any baseline differences (see Table 1). The number of trials needed to reach 85% criterion on Day 1 was subjected to a one-way between-subjects ANOVA, with the group (reminder, no reminder, control) as the independent variable. There were no significant differences between groups in the number of blocks to criterion, *F*(2, 63) = .12, *p* = .884, ηp2 = .004, BF10 = .14. In addition, a two-tailed *t* test was used to compare the reminder group to the no-reminder group for Day 2 acquisition. There was no significant difference in the number of learning trials to reach criterion for Day 2 learning, *t*(42) = −1.39, *p* = .173, *d* = −.42, BF10 = .64.

Likewise for Experiment 2 (2-day conditions; see Table 2), there was no significant difference between groups to reach criterion, *F*(2, 63) = .60, *p* = .554, ηp2 = .02, BF10 = .11. In addition, a one-tailed *t* test was used to compare the reminder group with the no-reminder group for Day 2 acquisition, which revealed no significant difference between the groups, *t*(42) = −.60, *p* = .550, *d* = −.28, BF10 = .34.

**Accuracy of free recall on Day 3** Mean accuracy of List-1 free recall is illustrated in Fig. 1 (mean accuracy for each of the four test trials is in Supplementary Table 1). To replicate Hupbach et al. (2007), we conducted a 3 × 4 mixed-design ANOVA, with group (reminder, no reminder, control) as the between-subjects factor and recall trial (1–4) as the within-subjects factor. For Experiment 1 (the 3-day conditions) there was a significant main effect of recall trial, *F*(1.56, 61) = 42.32, *p* < .001, ηp2 = .40. There was also a main effect of group, *F*(2, 63) = 9.16, *p* < .001, ηp2 = .23; unlike Hupbach et al., 2007). The interaction between recall trial and group was not significant, *F*(3.12, 122) = 2.35, *p* = .075, ηp2 = .07. This pattern of results was supported by Bayesian analyses; the best model included both main effects of recall trial and group, BF10 = 1.96E+19. We conducted

| Table 1 | Experiment 1. Mean trials (and standard deviations) taken to reach criterion |
|---------|-----------------------------|
| Condition | Day 1 | Day 2 |
| Reminder | 2.32 (.57) | 1.86 (.47) |
| No reminder | 2.41 (.67) | 2.09 (.61) |
| Control | 2.36 (.58) |

Note. *N* = 66

| Table 2 | Experiment 2. Mean trials (and standard deviations) taken to reach criterion |
|---------|-----------------------------|
| Condition | Day 1 | Day 2 |
| Reminder | 2.45 (.74) | 2.09 (.61) |
| No reminder | 2.36 (.66) | 2.23 (.87) |
| Control | 2.22 (.69) |

Note. *N* = 66
within-subject contrasts to determine whether there was a linear trend across trials. Like Hupbach et al. (2007), we found that participants recalled significantly more objects on each subsequent test, all ps < .05. We used two-tailed t tests to compare the reminder group with the no-reminder group and the no-reminder group to the control group. We found that the no-reminder group recalled significantly more correct objects compared with the reminder group, \( t(42) = -2.04, p = .047, d = -.62, BF_{10} = 1.54 \). The control group recalled significantly more objects on average compared with the no-reminder group, \( t(42) = -2.36, p = .023, d = -.71, BF_{10} = 2.62 \). With the frequentist statistics, these findings differ from Hupbach et al. (2007), who found only a main effect of trial, with a linear trend indicating increased recall accuracy across test trials. However, the small BF s found here suggest only anecdotal evidence for these differences.

For Experiment 2 (the 2-day conditions; see Supplementary Table 2), there was a significant main effect of trial with participants recalling significantly more correct objects on each subsequent test, \( F(1.29, 61) = 61.19, p < .001, \eta^2_p = .49 \). There was a main effect of group (unlike Hupbach et al., 2007), \( F(2, 63) = 5.11, p = .009, \eta^2_p = .14 \). The interaction between trial and group was not significant, \( F(2.60, 122) = .99, p = .392, \eta^2_p = .03 \). This pattern of results was supported by Bayesian analyses; the best model included a main effect of trial and group, \( BF_{10} = 6.34E+25 \). We followed this with two-tailed t tests to compare the reminder group with the no-reminder group and the no-reminder group to the control group. There was no significant difference between the no-reminder group and the reminder group, \( t(42) = -.82, p = .420, d = -.25, BF_{10} = .39 \). The control group recalled more objects than the no-reminder group, \( t(42) = -2.59, p = .013, d = -.78, BF_{10} = 3.98 \), although the BF suggests only moderate evidence for a difference.

**Intrusion errors on Day 1 recall** The primary measure of interest is the number of intrusions from Day 2 into memory of Day-1 objects (Fig. 2). Intrusions were analyzed with a 3 (group) \( \times \) 4 (trial) mixed-design ANOVA, with trial as the within-subjects factor. For Experiment 1, we found a significant main effect of trial, \( F(2.5, 63) = 4.57, p = .007, \eta^2_p = .07 \), and of group, \( F(2, 63) = 15.29, p < .001, \eta^2_p = .33 \). The interaction between trial and group was also significant, \( F(5, 122) = 2.47, p = .034, \eta^2_p = .07 \), indicating differences in intrusion rate across the groups. This pattern of results was supported by Bayesian analyses; the best model included both main effects and the interaction, \( BF_{10} = 22400.4 \). We used one-tailed t tests to compare the reminder group and the no-reminder group, and the no-reminder group to the control group. The reminder group showed significantly more List 2 intrusions compared with the no-reminder group, \( t(42) = 3.34, p < .001, d = 1.01, BF_{10} = 39.33 \). The no-reminder group showed significantly more List 2 intrusions compared with the control group, \( t(42) = 3.98, p < .001, d = 1.20, BF_{10} = 187.53 \). These findings differ from Hupbach et al., (2007), who found no significant differences between the no-reminder group and the control group.

---

1 This difference was not significant when outliers were removed, \( t(41) = -1.81, p = .08 \).
For Experiment 2 (the 2-day conditions), the main effect of trial was significant, $F(1.97, 63) = 4.75, p = .011, \eta_p^2 = .07$. Neither the main effect of group, $F(2, 63) = 1.64, p = .202, \eta_p^2 = .05$, nor the interaction between group and trial, $F(3.94, 63) = .91, p = .461, \eta_p^2 = .03$, were significant. Bayesian analyses supported this pattern; the best model included only the main effect of trial, $BF_{10} = 7.18$. One-tailed $t$ tests showed that there was no significant difference between the reminder group and the no-reminder group, $t(42) = .86, p = .198, d = .26, BF_{10} = .18$, nor between the no-reminder group and the control group, $t(42) = 1.32, p = .26, d = .33, BF_{10} = .48$. Thus, the null effect in the 2-day conditions found by Hupbach et al. (2007) was replicated.

Across experiment analysis In addition to Hupbach et al.’s (2007) analyses, we conducted a $2 \times 3$ (group) between-subjects ANOVA on the number of intrusions across test trials (see Fig. 2), in order to directly compare the 2-day and 3-day conditions. There was a significant main effect of group, $F(2, 131) = 16.68, p < .001, \eta_p^2 = .21$, and a significant main effect of experiment, $F(1, 2) = 20.22, p < .001, \eta_p^2 = .14$, and a significant interaction between the two, $F(2, 131) = 12.73, p < .001, \eta_p^2 = .17$. This pattern was supported by Bayesian analyses; the best model included both main effects and the interaction, $BF_{10} = 7592.76$. Follow-up $t$ tests showed that the reminder group in the 3-day condition recalled significantly more List 2 intrusions compared with the reminder group in the 2-day condition, $t(42) = -3.97, p < .001, d = -1.20, BF_{10} = 181.17$. The no-reminder group in the 3-day condition recalled significantly more List 2 intrusions compared with the no-reminder group in the 2-day condition, $t(42) = -2.94, p = .005, d = -.89, BF_{10} = 15.88$. There were no significant differences in List 2 intrusions between the 3-day and 2-day control groups, $t(42) = 1.09, p = .281, d = .33, BF_{10} = .48$.

Discussion

In this study, we investigated the time-dependency assumption of reconsolidation theory by directly replicating the experiments of Hupbach et al. (2007), who found that memory updating only occurred when there was time for a reconsolidation process to unfold. We replicated this finding: in the 3-day condition there was a significant and credible difference between the reminder groups for number of List-2 intrusions into Day 1 memory, but only null effects were found in the 2-day condition. Thus, we supported the time-dependency pattern that is critical to reconsolidation theory.

This pattern of results poses challenges to other accounts of memory updating. For instance, the memory integration theory argues that there is no need for a unique biological process (i.e., reconsolidation) to explain how older memories can be updated with new information (Gisquet-Verrier & Riccio, 2018, 2019). This account holds that memory updating is due to the integration of new information into a reactivated memory in a state-dependent manner. By this explanation, memory updating (or memory stability) should only be seen when the state of the subject at the final test matches the state of the subject during the reactivation phase. If there is a mismatch, memory updating will not be found. Evidence in support of this idea has been found in nonhuman animals (Gisquet-Verrier et al., 2015) and humans (Kiley & Parks, 2022). Despite that support, the current data are challenging because one tenet of the integration account is that memory updating is a rapid process that does not require a delay between Day 2 reactivation and the final test (see Bridge & Voss, 2014, for an example in a different paradigm). The current results would seem to disconfirm that hypothesis.

However, other researchers have argued that the results in the original study and this replication are not due to time dependency at all. Sederberg et al. (2011) argued that participants use a recall-to-reject strategy in the 2-day condition. Specifically, one major cue for recall is the temporal context, and participants can use that context to reject or accept items; if the context is “studied today,” they can be confident in rejecting the item. It could also be the case that List-2 items do not intrude very often in the 2-day condition because they can be rejected based on their greater strength relative to List-1 items. Thus, both specific retrieval of context and an assessment of memory strength could lead to the rejection of List 2 items in the 2-day condition such that there would be no difference between the reminder and no-reminder conditions. Ideally, one would eliminate the temporal context cue and the differences in memory strength between the two lists to test this idea. However, since the temporal context cue cannot be eliminated from this paradigm, one way to test this explanation would be to manipulate the strength of the lists such that they are closer to equal across Days 1 and 2 and/or to undermine the ability to use recall-to-reject by taxing the resources necessary to do so.

This recall-to-reject explanation could pose a strong challenge to the reconsolidation hypothesis. If it were the case that we see no evidence of memory updating in the 2-day condition because of differential memory strength and/or temporal context between the two lists that can be used as a source attribution cue, then we would not need a reconsolidation explanation to account for the results. However, the current data do support the reconsolidation explanation, and other accounts of the time-dependent pattern will need to be pitted against reconsolidation theory to determine if they can explain memory updating patterns better than reconsolidation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-022-02102-3.
References

Bridge, D. J., & Voss, J. L. (2014). Active retrieval facilitates across-episode binding by modulating the content of memory. *Neuropsychologia, 63*, 154–164.

Capelo, A. M., Albuquerque, P. B., & Cadavid, S. (2018). Exploring the role of context on the existing evidence for reconsolidation of episodic memory. *Memory, 27*(3), 280–294.

Elsey, J. W., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological Bulletin, 144*(8), 797.

Gisquet-Verrier, P., Lynch, J. F., Cutolo, P., Toledano, D., Ulmen, A., Jasnow, A. M., & Riccio, D. C. (2015). Integration of new information with active memory accounts for retrograde amnesia: A challenge to the consolidation/reconsolidation hypothesis? *Journal of Neuroscience, 35*(33), 11623–11633.

Gisquet-Verrier, P., & Riccio, D. C. (2018). Memory integration: An alternative to the consolidation/reconsolidation hypothesis. *Progress in Neurobiology, 171*, 15–31.

Gisquet-Verrier, P., & Riccio, D. C. (2019). Memory integration as a challenge to the consolidation/reconsolidation hypothesis: similarities, differences and perspectives. *Frontiers in Systems Neuroscience, 12*, 71.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory, 14*, 47–53.

Hupbach, A., Gomez, R., & Nadel, L. (2011). Episodic memory updating: The role of context familiarity. *Psychonomic Bulletin & Review, 18*, 787–797.

Inda, M. C., Muravieva, E. V., & Alberini, C. M. (2011). Memory retrieval and the passage of time: From reconsolidation and strengthening to extinction. *Journal of Neuroscience, 31*(5), 1635–1643.

Kiley, C., & Parks, C. M. (2022). Mechanisms of memory updating: State dependency vs. reconsolidation. *Journal of Cognition, 5*, 7.

Klingmüller, A., Caplan, J. B., & Sommer, T. (2017). Intrusions in episodic memory: Reconsolidation or interference? *Learning & Memory, 24*(5), 216–224.

Scully, I. D., Napper, L. E., & Hupbach, A. (2017). Does reactivation trigger episodic memory change? A meta-analysis. *Neurobiology of Learning and Memory, 142*, 99–107.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review, 18*(3), 455–468.

Open practices statement The data and protocol for these experiments are available on the Open Science Framework (https://osf.io/zr7k2/).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.