Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning

Sevenster, D.; Beckers, T.; Kindt, M.

Link to publication

Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning

Dieuwke Sevenster,[1,2] Tom Beckers,[1,2,3] and Merel Kindt[1,2]

[1]Department of Clinical Psychology, University of Amsterdam, 1018 WB Amsterdam, The Netherlands; [2]Amsterdam Brain and Cognition Center, University of Amsterdam, 1018 WS Amsterdam, The Netherlands; [3]Department of Psychology, University of Leuven, B-3000 Leuven, Belgium

Although disrupting reconsolidation is promising in targeting emotional memories, the conditions under which memory becomes labile are still unclear. The current study showed that post-retrieval changes in expectancy as an index for prediction error may serve as a read-out for the underlying processes engaged by memory reactivation. Minor environmental changes define whether retrieval induces memory reconsolidation or the initiation of a new memory trace even before fear extinction can be observed.

[Supplemental material is available for this article.]

The recent appreciation of memory as a dynamic rather than static process inspires new fundamental questions about memory malleability. After the formation of associative fear memory, a re-encounter with the feared stimulus can lead to various outcomes; retrieval only, reconsolidation, or new learning. Insights into how the system determines whether a memory trace should be updated or an additional memory trace should be formed are essential not only for our understanding of the fundamentals of learning and memory but also for the development of reconsolidation based treatment.

Prediction error (PE)—a mismatch between what is expected based on previous experiences and the actual state of events (Rescorla and Wagner 1972)—is a necessary condition for inducing reconsolidation of human associative fear memory upon retrieval (Sevenster et al. 2013). Previously, we have shown the utility of retrieval-induced changes in the expectancy ratings of the unconditioned stimulus (US) during presentation of the conditioned stimulus (CS) as a measure of PE (Sevenster et al. 2013). This behavioral and noninvasive index of PE may indicate whether the original memory trace was destabilized by the retrieval experience, independent of the outcome of the reconsolidation process itself. It may, however, be questioned whether PE is a sufficient condition for reconsolidation, given that PE can also give rise to new learning instead (e.g., extinction learning) (Rescorla and Wagner 1972). For example, extinction training, which involves the formation of a new inhibitory memory trace (Bouton 2002), puts a constraint on reconsolidation (Eisenberg et al. 2003; Lee et al. 2006; Bos et al. 2012). During extinction training, repeated or prolonged unreinforced exposure generates multiple PEs, which eventually reduces both threat expectancy and fear responding. However, the transition from updating of the original memory trace to the formation of a new memory trace may occur long before the expression of the inhibitory extinction memory can be observed. Until now, the restraint that new learning puts on reconsolidation could only be inferred from extinction of the fear behavior itself. In the current study, we tested whether changes in threat expectancy could serve not only as an index of memory destabilization, but could also reveal a boundary condition of reconsolidation independent from the fear reduction itself. Hereto, we established the optimal parameters for fear acquisition and memory reactivation that lead to memory destabilization. From there, we increased the number of reactivation trials, without already extinguishing the fear response. We tested two alternative hypotheses regarding the optimal conditions for inducing memory reconsolidation: (1) PE is necessary, but not sufficient: Reconsolidation is triggered by a single PE but does not take place when memory reactivation induces multiple PEs even if the fear expression itself is not reduced or (2) PE is necessary and sufficient: Fear reconsolidation is triggered by a single or multiple PEs as long as no change in the target behavior (i.e., extinction of the fear response) is observed during memory retrieval.

In a human differential fear conditioning paradigm, we created three groups in which fear acquisition was partially reinforced (50%; reinforcement on all even but not on odd trials). Participants were instructed that one of the two pictures (CS1) was followed by a shock (US) on 50% of the trials, while the other was not (CS2). One day later the memory was reactivated with one (no PE, $n = 18$), two (single PE group, $n = 18$) or four unreinforced reminder trials (multiple PEs group, $n = 16$), followed by administration of the noradrenergic β-blocker propranolol (40 mg). Thus, participants either had no opportunity (no PE), a single (single PE) or multiple opportunities (multiple PEs) to change their expectations. Given the reinforcement schedule (50%) on Day 1, a single reminder trial should not generate PE-driven cognitive learning since there is a 50% chance that the CS will be reinforced (or not) (no PE group). In contrast, two unreinforced reminder trials should induce a single PE and subsequent reconsolidation (single PE group), since reinforcement—that was expected on the second trial due to the 50% reinforced schedule during acquisition—stayed off. The crucial test, however, was whether reconsolidation would be either prevented or triggered by four nonreinforced retrieval trials (multiple PEs group). We employed potentiation of the startle response as a measure of fear responding, while

Corresponding author: m.kindt@uva.nl
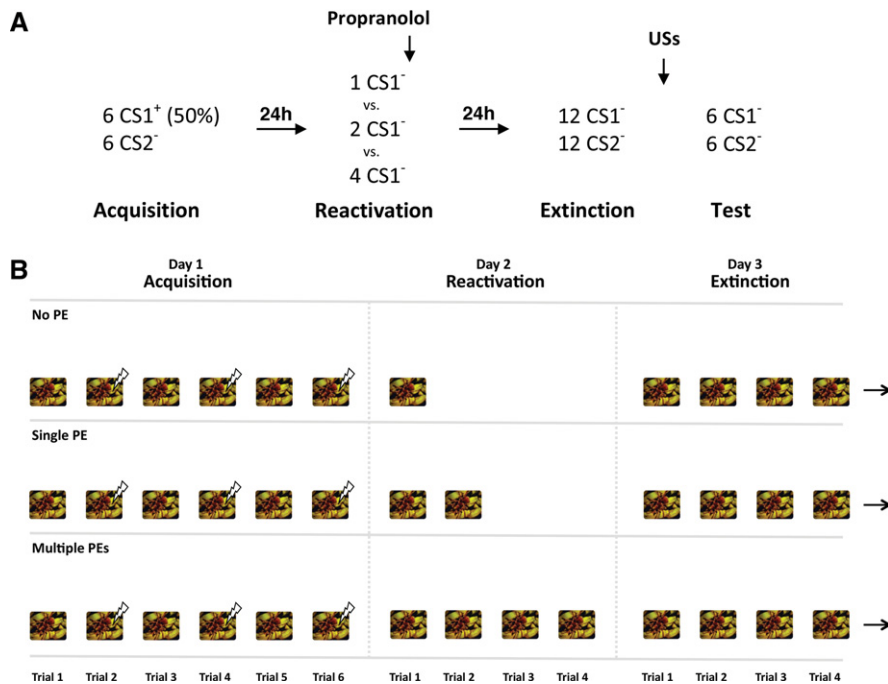
Figure 1. (*A*) Schematic representation of the experimental design. (*B*) Reinforcement schedule of the CS1 during the acquisition, reactivation and the first trials of the extinction phase for the experimental groups. CS1 is depicted as one of the two images used as CSs. US (electrical stimulus) is depicted as a lightning bolt. Given that participants learned that every other CS1 was followed by shock (50% reinforcement schedule), a single reactivation trial (no PE group) should not induce PE. A second unreinforced reactivation trial should induce PE (single PE group) and four unreinforced reactivation trials should generate multiple PEs (multiple PE group).

declarative US-expectancy ratings served to index PE. On Day 3, all groups underwent an extinction and reinstatement procedure to test the absence of fear memory expression (Fig. 1A,B). We expected that noradrenergic blockade would disrupt reconsolidation, evidenced by a reduction in differential conditioned startle fear responding 1 d after memory reactivation in the single PE group, while propranolol should not affect the startle response in the no PE group. For the multiple PE group we tested two opposing hypotheses. If the four reminder trials are already a boundary condition for memory reconsolidation (multiple PEs group), the noradrenergic β-blocker propranolol should not affect the fear expression. On the other hand, if the transition from reconsolidation to extinction would only take place after successful extinction training (i.e., significant decline in fear responding), propranolol should still interfere with the process of memory reconsolidation after four reminder trials, thereby reducing the fear response 1 d later.

We observed acquisition of differential US-expectancies (Fig. 2) and startle potentiation (Fig. 3) on Day 1 in all groups. Acquisition of US-expectancies and startle responding did not differ between the groups (see Supplemental Material).
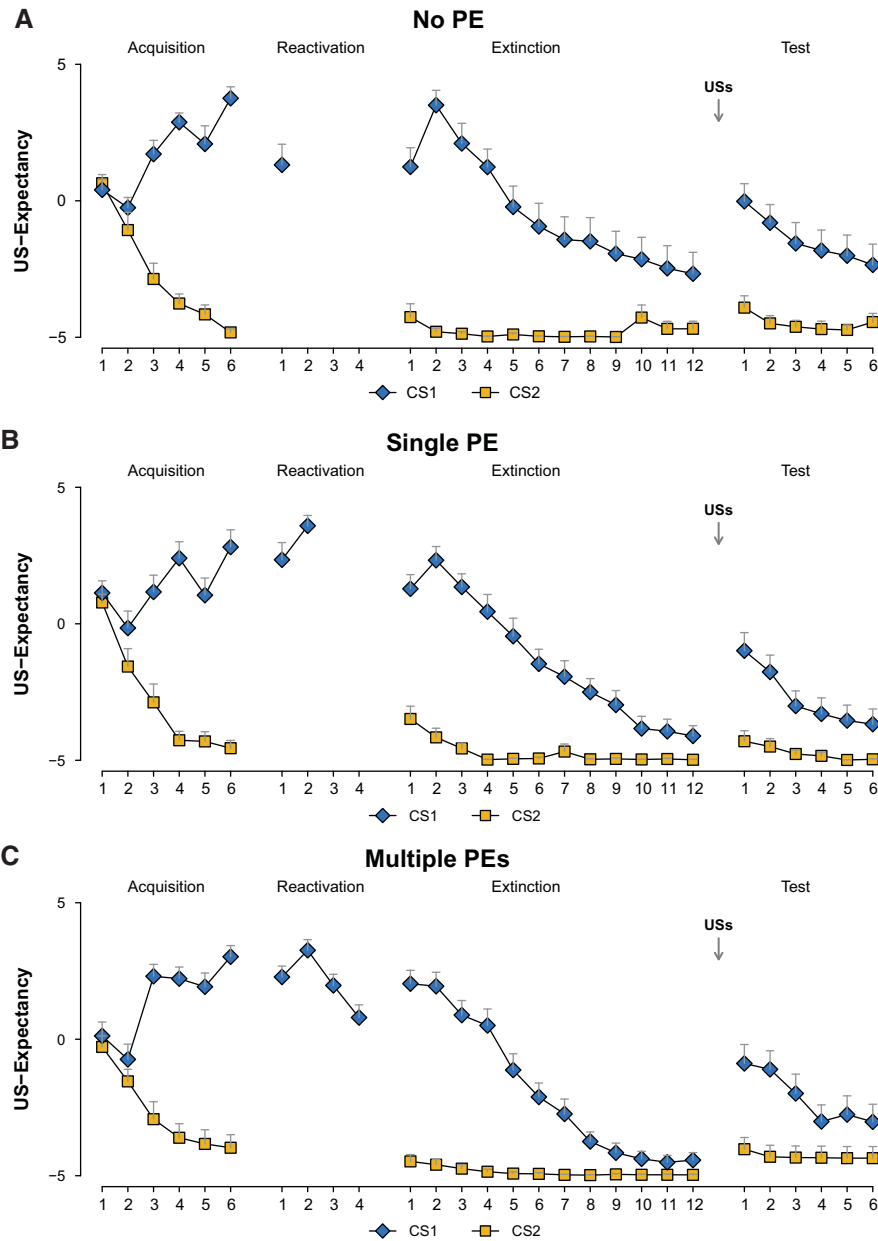
To investigate the effectiveness of our manipulation (i.e., number of reactivation trials) in inducing PE, we analyzed the pattern of differential US-expectancies on Day 3. Given the 50% reinforcement schedule of Day 1, we expected that the absence of reinforcement on Day 2 would affect differential responding on the second trial but not so much on the first trial of Day 3. Also, we expected that the number of reactivation trials would differentially affect US-expectancies on Day 3. In line with our expectation, the number of reactivation trials (1 vs. 2 vs. 4) differentially affected shock expectancy on the first two extinction trials of Day 3 (stimulus × trial × group, $F_{(2,49)} = 4.26$, $P < 0.02$, $\eta^2_p =$

0.15). Importantly, the groups near-significantly differed in US-expectancy ratings (CS1 vs. CS2) on the second (stimulus × group, $F_{(2,49)} = 3.04$, $P < 0.057$, $\eta^2_p = 0.11$, $1 - \beta = 0.56$) but not the first (stimulus × group; $F_{(2,49)} < 1.24$) extinction trial. Differential ratings on the second trial of extinction were greater for the no PE group relative to the single PE (stimulus × group, $F_{(1,34)} = 4.37$, $P < 0.044$, $\eta^2_p = 0.11$) and the multiple PEs group (stimulus × group, $F_{(1,32)} = 4.39$, $P < 0.044$, $\eta^2_p = 0.12$). Hence, the reactivation procedures differentially affected US-expectancy ratings on the second but not the first trial of extinction.

Reduction of differential US-expectancies during extinction training on Day 3 also differed between the groups (trial 1 vs. 12; trial × group; $F_{(2,49)} = 3.32$, $P < 0.044$, $\eta^2_p = 0.12$). There was more reduction in the multiple PEs group compared with the no PE group (trial × group; $F_{(1,32)} = 4.85$, $P < 0.035$, $\eta^2_p = 0.13$) but there was no stronger reduction in the multiple PEs group relative to the single PE group (trial × group; $F_{(1,32)} < 1$). Finally, there was a near-significant greater reduction in expectancy ratings during extinction in the single PE group compared with the no PE group (trial × group; $F_{(1,34)} = 3.91$, $P < 0.056$, $\eta^2_p = 0.10$, $1 - \beta = 0.49$). Given the increase in US-expectancies from the first to the second extinction trial in the no PE and single PE group, we additionally analyzed extinction from the second to the last extinction trial and observed similar differences between groups (see Supplemental Material). For analyses of effects of reinstatement on US-expectancy see Supplemental Material. In conclusion, presenting four unreinforced reminder trials affected extinction learning on Day 3, while a single reminder trial did not generate any new learning. Note that by design the single PE group holds an intermediate position. The observation of more extinction on Day 3 in the single PE group relative to the no PE group suggests that two unreinforced trials induced more learning relative to the no PE group the day before. Also, more extinction on Day 3 in the multiple PEs group relative to the single PE group suggests that the four unreinforced trials presented on the previous day resulted in most PE-driven learning. The groups did not differ in US-expectancy ratings during the first trial of reactivation ($F_{(2,51)} < 1$). While there was an initial increase in expectancy ratings from the first to the second trial of reactivation in the single PE and multiple PEs groups (main effect trial, $F_{(1,32)} = 9.11$, $P < 0.005$, $\eta^2_p = 0.22$; trial × group, $F < 1$), a reduction of US-expectancy ratings during reactivation was eventually observed in the multiple PEs group (trials 1–4; multiple PEs group) (main effect trial, $F_{(1.93,28.93)} = 8.5$, $P < 0.001$, $\eta^2_p = 0.36$).

Whereas US-expectancy was reduced during memory reactivation in the multiple PEs group, fear expression during reactivation remained intact in all groups (see Supplemental Material). Yet, the groups differed on the fear retention test one day later (stimulus × group, $F_{(2,49)} = 3.32$, $P < 0.044$, $\eta^2_p = 0.12$). As expected, propranolol completely eliminated differential startle potentiation (CS1 vs. CS2) on the first trial of extinction in the single PE group (main effect stimulus, $F_{(1,17)} < 1$) (Fig. 3B). In contrast, the fear response was still intact not only when no prediction error

A No PE

B Single PE

C Multiple PEs

Figure 2. Online US-expectancy ratings as a measure of PE can indicate the transition from retrieval, to reconsolidation, to new learning. Mean US-expectancy ratings to the CS1 and CS2 trials during acquisition, reactivation, extinction and reinstatement test for the No PE group ($n = 18$, A), the Single PE group ($n = 18$, B), and the Multiple PEs group ($n = 16$, C). Error bars represent SEM.

significance with a moderate effect size (stimulus × group, $F_{(2,49)} = 2.96$, $P < 0.061$, $\eta^2_p = 0.11$, $1 - \beta = 0.55$). Crucially, the unexpected USs did not reinstate the startle fear response in the single PE group (Fig. 3B), evidenced by an absence of differential startle responding on the test trial (main effect stimulus, $F_{(1,17)} < 1$). Follow-up analyses indeed revealed significantly less differential responding in the single PE group compared with both the no PE (stimulus × group, $F_{(1,34)} = 4.06$, $P < 0.05$, $\eta^2_p = 0.11$) and the multiple PEs group (stimulus × group, $F_{(1,32)} = 4.83$, $P < 0.035$, $\eta^2_p = 0.13$). A return of fear was observed in both the no PE (Fig. 3A) and multiple PEs group (Fig. 3C), evidenced by significantly more startle responding to the CS1 compared with the CS2 (main effect stimulus, $F_{(1,32)} = 10.15$, $P < 0.003$, $\eta^2_p = 0.24$; stimulus × group, $F_{(1,32)} < 1$) (Fig. 3). Thus, noradrenergic blockade neither eliminated fear directly nor prevented the return of fear after successful extinction when there was no PE at all or when multiple PEs reduced US-expectancy during memory retrieval. In line with our previous studies, propranolol did not affect the skin conductance response (see Supplemental Material; Soeter and Kindt 2010, 2011a,b, 2012; Sevenster et al. 2012).

We showed that memory reactivation induces simple retrieval, reconsolidation or new learning (Osan et al. 2011; Piñeyro et al. 2014), as a function of the number of reactivation trials. Our data showed that in the absence of PE, the retrieved memory remained intact, although a single PE was required for memory destabilization. Reconsolidation was, however, no longer triggered when memory reactivation induced too much new learning.

In line with our previous findings we provided evidence that PE is indeed a necessary condition for reconsolidation (Sevenster et al. 2012, 2013), but not a sufficient condition to trigger memory reconsolidation. Given that propranolol did not reduce fear responding in the multiple PEs group, it may be concluded that only four unreinforced trials during memory reactivation prevented reconsolidation. It was already known that extinction of fear during memory retrieval could operate as a boundary condition for memory reconsolidation (Eisenberg et al. 2003; Bos et al. 2012). Here we demonstrate that reconsolidation can also be prevented long before the fear expression itself is extinguished and that a change in US-expectancy ratings may serve to indicate this boundary condition for memory reconsolidation. Thus, from the fear expression itself we cannot infer the underlying process that is ultimately engaged by memory reactivation (see also Mamou et al. 2006). In contrast, US-expectancy is sensitive to relatively minor changes during memory reactivation, which may not only unveil the switch from stability to plasticity but also

was generated during memory retrieval (no PE) (Fig. 3A), but also when sufficient PE was generated to induce immediate change in declarative US expectancies (multiple PEs-group) (Fig. 3C), evidenced by higher startle responding to the CS1 compared with the CS2 in the no PE and multiple PEs group (main effect stimulus, $F_{(1,32)} = 17.46$, $P < 0.001$, $\eta^2_p = 0.35$; stimulus × group, $F_{(1,33)} < 1$). Given that a difference in startle response already appeared on the first trial of extinction, the groups differed over the course of extinction training (Trials 1–12) (stimulus × group, $F_{(2,49)} = 3.53$, $P < 0.037$, $\eta^2_p = 0.13$) (follow-up extinction analyses are available in Supplemental Material).

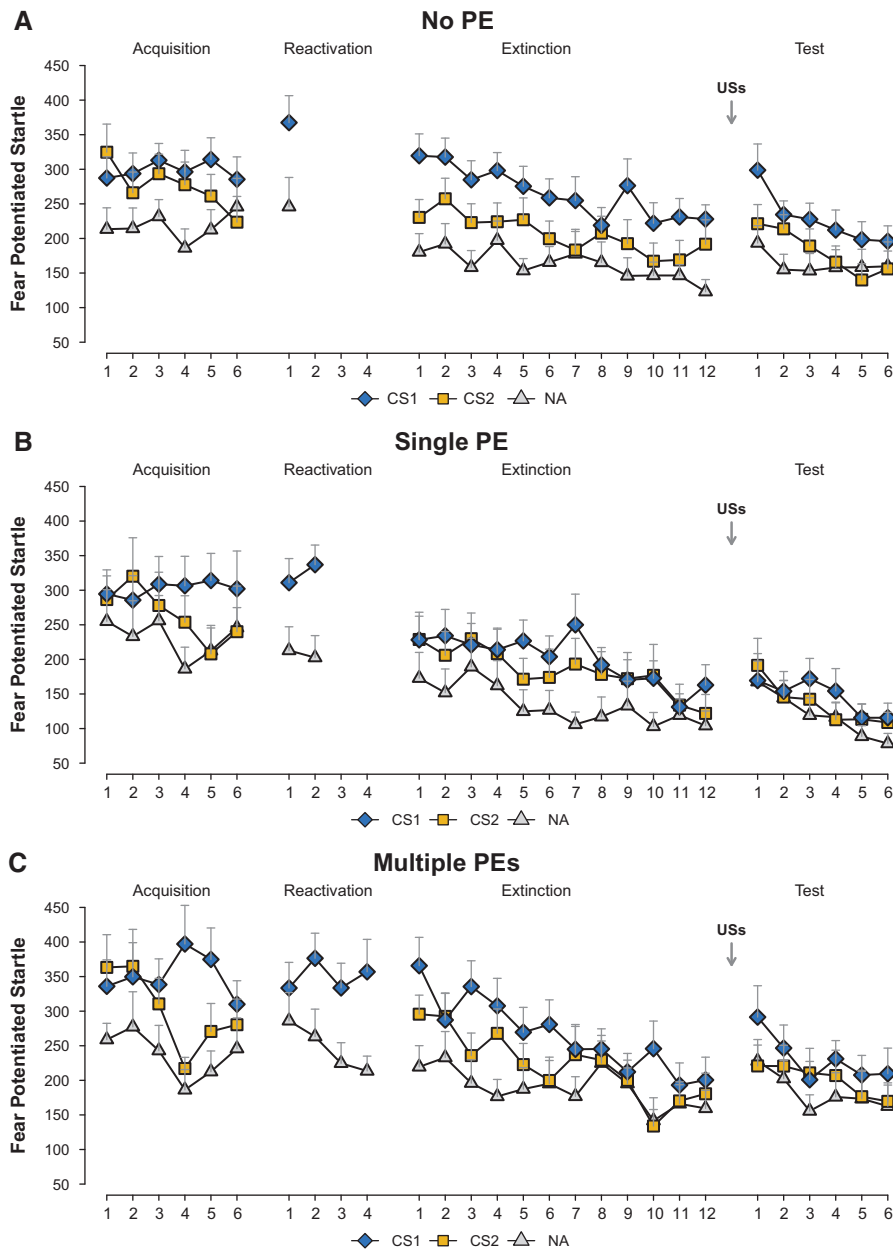The difference in startle fear responding (CS1 vs. CS2) between the three groups on the reinstatement test trial approached

Figure 3. PE is a necessary but not sufficient condition for reconsolidation. Mean startle responses to the CS1, CS2, NA trials during acquisition, reactivation, extinction, and reinstatement test (*A,C*). Propranolol affected the startle response only in the Single PE (*n* = 18, *B*), but neither in the No PE group (*n* = 18, *A*) nor in the Multiple PEs group (*n* = 16, *C*). Error bars represent SEM.

the transition from plasticity to the formation of a new memory trace. From a clinical point of view, it is very promising that reduction of US-expectancy may indicate that memory reactivation no longer engages reconsolidation and, hence, that pharmacological manipulation aimed to disrupt reconsolidation is no longer effective.

In our previous studies, it became clear that noradrenergic blockade affects the fear potentiated startle while the US-expectancy ratings generally remained unaffected (Kindt et al. 2009; Soeter and Kindt 2010; Sevenster et al. 2013). Only in one study, where propranolol was administered before extinction training, the noradrenergic beta-blocker actually impeded the US-expec-

tancy ratings but not the startle response. This effect persisted one day later on both the extinction retention and reinstatement test (Bos et al. 2012). The direct and delayed effects of propranolol on US-expectancy ratings were not observed on the first trials of extinction and re-extinction but only after multiple unreinforced reminder presentations. This suggests that the noradrenergic beta-blocker affected the expression of declarative memory only when threat expectancies were adjusted to the consistent absence of threat.

We have no evidence that presentation of four unreinforced reminder trials (multiple PEs) already engaged fear extinction. Interestingly, recent studies in rodents (Flavell and Lee 2013; Merlo et al. 2014) provide compelling evidence that our multiple PEs protocol induced neither reconsolidation nor extinction. Bidirectional manipulations of NMDAr activity differentially affected reconsolidation and extinction when applied to a short or long reactivation protocol, respectively (Merlo et al. 2014). In marked contrast, the animals that received an intermediate reactivation protocol were completely insensitive to NMDAr manipulations. It is therefore likely that the four trial reactivation protocol in the current study induced a similar transitional state between reconsolidation and extinction. Thus, reconsolidation and extinction are not only mutually exclusive but seem also to be separated by a transitional phase during which neither of the two opposite processes is engaged.

In sum, the window of opportunity to target emotional memory with amnesic agents is small, since it is preceded and followed by phases that leave the original memory unaffected. Even though the fear reducing effects are very robust and promising for the development of reconsolidation-based treatments, the success of the manipulation depends on subtle differences in the reactivation procedure. This poses a real challenge for clinical practice, which can be resolved by careful selection of the reactivation parameters. Changes in declarative knowledge (US-expectancy) can serve as an independent index of PE that may help to discern whether an intervention to change maladaptive emotional memories has the potential to be effective.

Acknowledgments

References

Bos MGN, Beckers T, Kindt M. 2012. The effects of noradrenergic blockade on extinction in humans. *Biol Psychol* **89:** 598–605.

Bouton ME. 2002. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol Psychiatry* **52:** 976–986.

Eisenberg M, Kobilo T, Berman DE, Dudai Y. 2003. Stability of retrieved memory: inverse correlation with trace dominance. *Science* **301:** 1102–1104.

Flavell CR, Lee JLC. 2013. Reconsolidation and extinction of an appetitive Pavlovian memory. *Neurobiol Learn Mem* **104:** 25–31.

Kindt M, Soeter M, Vervliet B. 2009. Beyond extinction: erasing human fear responses and preventing the return of fear. *Nat Neurosci* **12:** 256–258.

Lee JLC, Milton AL, Everitt BJ. 2006. Reconsolidation and extinction of conditioned fear: inhibition and potentiation. *J Neurosci* **26:** 10051–10056.

Mamou CB, Gamache K, Nader K. 2006. NMDA receptors are critical for unleashing consolidated auditory fear memories. *Nat Neurosci* **9:** 1237–1239.

Merlo E, Milton AL, Goozée ZY, Theobald DE, Everitt BJ. 2014. Reconsolidation and extinction are dissociable and mutually exclusive processes: behavioral and molecular evidence. *J Neurosci* **34:** 2422–2431.

Osan R, Tort AB, Amaral OB. 2011. A mismatch-based model for memory reconsolidation and extinction in attractor networks. *PLoS One* **6:** e23113.

Piñeyro ME, Monti RIF, Alfei JM, Bueno AM, Urcelay GP. 2014. Memory destabilization is critical for the success of the reactivation–extinction procedure. *Learn Mem* **21:** 46–54.

Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: Current theory and research*, pp. 64–99. Appleton-Century-Crofts, New York.

Sevenster D, Beckers T, Kindt M. 2012. Retrieval per se is not sufficient to trigger reconsolidation of human fear memory. *Neurobiol Learn Mem* **97:** 338–345.

Sevenster D, Beckers T, Kindt M. 2013. Prediction error governs pharmacologically induced amnesia for learned fear. *Science* **339:** 830–833.

Soeter M, Kindt M. 2010. Dissociating response systems: erasing fear from memory. *Neurobiol Learn Mem* **94:** 30–41.

Soeter M, Kindt M. 2011a. Disrupting reconsolidation: pharmacological and behavioral manipulations. *Learn Mem* **18:** 357–366.

Soeter M, Kindt M. 2011b. Stimulation of the noradrenergic system during memory formation impairs extinction learning but not the disruption of reconsolidation. *Neuropsychopharmacology* **37:** 1204–1215.

Soeter M, Kindt M. 2012. Erasing fear for an imagined threat event. *Psychoneuroendocrinology* **37:** 1769–1779.