Fear memory uncovered: Prediction error as the key to memory plasticity
Sevenster, D.

Citation for published version (APA):
Sevenster, D. (2014). Fear memory uncovered: Prediction error as the key to memory plasticity
Chapter 5

Prediction error governs pharmacologically induced amnesia for learned fear

Dieuwke Sevenster
Tom Beckers
Merel Kindt

Science, 2013, 339: 830-833
ABSTRACT - Although reconsolidation opens up new avenues to erase excessive fear memory, subtle boundary conditions put constraints on retrieval-induced plasticity. Reconsolidation may only take place when memory reactivation involves an experience that engages new learning (prediction error). Thus far it has not been possible to determine the optimal degree of novelty required for destabilizing the memory. The occurrence of prediction error could only be inferred from the observation of a reconsolidation process itself. Here, we provide a non-invasive index of memory destabilization that is independent from the occurrence of reconsolidation. Using this index we show in humans that prediction error is (i) a necessary condition for reconsolidation of associative fear memory and (ii) determined by the interaction between original learning and retrieval. Insight into the process of memory updating is crucial for understanding the optimal and boundary conditions on reconsolidation and provides a clear guide for the development of reconsolidation-based treatments.
BACKGROUND

A consolidated fear memory can enter a transient labile phase upon its reactivation. Pharmacological blockade of the subsequent protein synthesis dependent restabilization (reconsolidation) produces a memory deficit in both animals (Nader et al., 2000) and humans (Kindt et al., 2009). However, an independent measure for memory destabilization, other than the occurrence of reconsolidation itself, is not yet available. The functional role of reconsolidation might be to keep memories up to date with new learning. Indeed, reconsolidation is triggered only when there is opportunity for new learning to take place during reactivation (Morris et al., 2006; Pedreira et al., 2004; Sevenster et al., 2012a).

Because associative learning requires prediction error (PE) (i.e., a discrepancy between actual and expected events) (Rescorla & Wagner, 1972), reconsolidation might also be a PE driven process. Even though it has frequently been suggested, there is no experimental evidence that PE is a necessary condition for reconsolidation. So far, PE could only be inferred from effective reconsolidation without an independent assessment of PE driven re-learning (Morris et al., 2006; Pedreira et al., 2004; Sevenster et al., 2012a). Unveiling a crucial role for PE in reconsolidation of fear memory — which may serve as an index for memory destabilization independent from the process of reconsolidation itself — will provide a clear guide for developing treatments to permanently reduce unwanted and excessive fears (such as posttraumatic stress disorder).

General associative learning models (Rescorla & Wagner, 1972) argue that PE is not determined by the mere co-occurrence of the conditioned stimulus (CS) and unconditioned stimulus (US), but by the discrepancy between what has already been learned (learning history) and what can be learned on a given trial. If memory retrieval follows a fully reinforced asymptotic learning episode, omission of a predicted reinforcement during reactivation (negative PE) (Waelti, Dickinson, & Schultz, 2001) may destabilize a consolidated memory during its reactivation, whilst a reinforced reactivation trial would leave the memory intact given that PE would then be absent. In contrast, if memory retrieval follows a partially reinforced, non-asymptotic learning episode, a similar reinforced reminder trial (positive PE) (Waelti et al., 2001) should generate additional learning and consequently be capable of inducing post-retrieval plasticity, because memory strengthening through further learning also requires reconsolidation mechanisms (Lee, 2008).

The noradrenergic β-blocker propranolol administered either before or after memory retrieval eliminates affective responding (fear potentiated startle) in
human participants but leaves the predominantly cognitive component of fear (US-expectancy ratings, skin conductance response) intact (Soeter & Kindt, 2010, 2011b). Given that propranolol does not affect declarative memory when reactivated with a single trial, online US-expectancy ratings during acquisition, retrieval and test could serve as an independent measure to test whether PE driven re-learning during reactivation is essential for reconsolidation of affective fear memory. The current study had a threefold aim: (i) to examine the role of PE in reconsolidation of fear memory, (ii) to examine whether PE depends on the interaction between the available information during reactivation and the learning history and (iii) to provide a measure for memory destabilization that is independent from the occurrence of reconsolidation itself.

In a human differential fear conditioning paradigm we tested two groups in which fear acquisition was fully reinforced (100% of the trials) (Fig. 1A,B). To ensure that asymptotic learning was indeed realized, the participants received explicit instructions regarding the contingencies between the conditioned stimuli (CS) and the unconditioned stimulus (US). On day 2 the memory was reactivated through either an unreinforced (Negative PE group; n = 15) or a reinforced (No PE group; n = 15) reactivation trial, followed by administration of propranolol (40 mg) (Fig. 1A,B). PE driven cognitive re-learning and corresponding reconsolidation should occur in the Negative PE group but not in the No PE group. We tested a third group to examine whether PE depends on the interaction between the information presented during reactivation and the learning history. In this group acquisition was partially reinforced (33% of the trials) and the memory was reactivated with a reinforced reminder trial (Positive PE group; n = 15), followed by administration of propranolol (40 mg) (Fig. 1A,B). In contrast to the full-reinforcement condition, here the reinforced reminder trial should induce PE driven additional learning given that a partial reinforcement schedule will not induce asymptotic learning. As such, a reinforced reactivation trial might also induce reconsolidation. Noradrenergic blockade after memory retrieval should disrupt reconsolidation — operationalized as a reduction in conditioned startle fear responding — in both the Negative PE and Positive PE group but not in the No PE group. On day 3 all groups underwent an extinction session followed by a reinstatement procedure to test to what extent the original fear memory trace was weakened (Fig. 1A,B).
RESULTS

All three groups showed fear learning and memory reactivation on day 1 and day 2 respectively for the startle fear response and online US-expectancy ratings (Supplementary Data). Analyses of differential US-expectancy ratings (CS1 vs. CS2) from the last trial of acquisition (day 1) to the first trial of extinction (day 3) revealed differences between the three groups (stimulus x trial x group, $F_{(2, 42)} = 25.44$, $p < .001$, $\eta^2_p = .55$). Follow-up analyses of the differential US-expectancy ratings from the last trial of acquisition (day 1) to the first trial of extinction (day 3) revealed a decrease in the Negative PE group (stimulus x trial x group, $F_{(1, 28)} = 8.18$, $p < .01$).
Figure 2. Mean US-expectancy ratings to the CS1 and CS2 trials during acquisition, reactivation, extinction and reinstatement test (A-C). US-expectancy ratings decreased from the end of acquisition to the beginning of extinction in the Negative PE group ($n = 15$, A), remained similar in the No PE group ($n = 15$, B) and increased in the Positive PE group ($n = 15$, C). Error bars represent SEM.

$p < .008, \eta^2 = .23$ and an increase in the Positive PE group (stimulus x trial x group, $F_{(1, 28)} = 42.98, p < .001, \eta^2 = .61$) relative to the No PE group (Fig. 2). A non-reinforced reactivation trial resulted in a decrease in differential US-expectancy ratings from the last trial of acquisition (day 1) to the first trial of extinction (day 3) when acquisition was fully reinforced (Negative PE) (stimulus x trial, $F_{(1, 14)} = 18.46, p$
Figure 3. Mean startle responses to the CS1 and CS2 trials during acquisition, reactivation, extinction and reinstatement test (A-C). Propranolol affected the startle response in both the Negative PE group (n = 15, A) and Positive PE group (n = 15, C) but not in the No PE group (n = 15, B). Error bars represent SEM.

< .001, $\eta^2_p = .57$) (Fig. 2A). Reinforcement of reactivation left the US-expectancy ratings unaffected in the No PE group (stimulus x trial, $F(1, 14) < 2.47$) (Fig. 2B). However, when acquisition had been partially reinforced a similar reactivation trial resulted in an increase in US-expectancy ratings (Positive PE group; stimulus x trial, $F(1, 14) = 31.72, p < .001, \eta^2_p = .69$) (Fig. 2C).
The three groups differed in startle responding on the first retention trial of extinction on day 3 (stimulus x group, $F_{(2, 42)} = 6.49, p < .003, \eta^2_p = .24$). Propranolol reduced the differential startle response (CS1 vs. CS2) on the first extinction trial in both the Negative PE (stimulus x group, $F_{(1, 28)} = 10.76, p < .003, \eta^2_p = .28$) and Positive PE group (stimulus x group, $F_{(1, 28)} = 7.89, p < .009, \eta^2_p = .22$), compared to the No PE group. Indeed, propranolol completely erased differential responding on the first extinction trial in the Negative PE (main effect stimulus, $F_{(1, 14) < 1}$) (Fig. 3A) and Positive PE group (main effect stimulus, $F_{(1, 14) < 1}$) (Fig. 3B). In contrast, this propranolol-induced amnesia was not observed when reactivation was devoid of new learning, as indicated by the differential startle response that was still present on the first extinction trial in the No PE group (main effect stimulus, $F_{(1, 14) = 13.47, p < .003, \eta^2_p = .49}$) (Fig. 3C). Given that propranolol eliminated differential responding in the reactivation conditions under which new learning occurred, the three groups differed over the course of extinction learning (trial 1 vs. trial 12) (stimulus x trial x group, $F_{(2, 42) = 5.23, p < .009, \eta^2_p = .20}$) (Supplementary Data). Thus, propranolol affected startle fear responding only in those groups (Negative PE and Positive PE) in which the reactivation trial resulted in changes in cognitive learning, be it incremental or decremental.

Differences in startle fear responding (CS1 vs. CS2) between the three groups on the reinstatement test trial approached significance (stimulus x group, $F_{(2, 42)} = 2.25, p < .118, \eta^2_p = .10$). This small effect can be attributed to a general increase in startle responding from the end of extinction (trial 12) to the test trial in the No PE group (main effect trial, $F_{(1, 14) = 10.20, p < .006, \eta^2_p = .42}$), which is typically observed after unpredictable shocks following fear extinction (2). Re-analysing the differential startle response to the test trial with the noise alone (NA) trial as the control stimulus (CS1 vs. NA) revealed, however, a significant difference between the three groups (stimulus x group, $F_{(2, 42)} = 5.57, p < .007, \eta^2_p = .21$). Follow-up analyses revealed significantly more differential responding in the No PE group compared to both the Negative PE (stimulus x group, $F_{(1, 28) = 13.25, p < .001, \eta^2_p = .32}$) and Positive PE group (stimulus x group, $F_{(1, 28) = 7.38, p < .011, \eta^2_p = .21}$). The startle response indeed recovered in the No PE group as indicated by stronger conditioned responding to the CS1 compared to the NA (main effect stimulus, $F_{(1, 14) = 24.01, p < .001, \eta^2_p = .63}$) (Fig. 3B), while no return of fear was observed in either the Negative PE (main effect stimulus, $F_{(1, 14) < 1.44}$) (Fig. 3A) or the Positive PE group (main effect stimulus, $F_{(1, 14) < 1}$) (Fig. 3C). Affective fear memory was only disrupted when actual learning took place during memory retrieval, showing that
post-retrieval plasticity depends on PE driven re-learning. Fear memory destabilization was not necessarily triggered by the absence of US-reinforcement (i.e., an extinction trial) but was also induced by a reinforced retrieval trial when fear learning on the previous day involved a partial reinforcement schedule. PE was determined by the interaction between the learning history and the retrieval session. PE driven learning — operationalized by a change in US-expectancy from the end of acquisition (day 1) to the beginning of memory testing (day 3)— may be used as a non-invasive index for memory destabilization.

**DISCUSSION**

The application of post-retrieval amnesic agents is considered to be a highly promising procedure to target excessive emotional memories typically observed in patients suffering from psychiatric disorders (such as posttraumatic stress disorder, addiction). However, the feasibility of disrupting reconsolidation may also be criticized given the subtle boundary conditions under which the amnesic agents do not affect memory (for review see Nader & Hardt, 2009). Reconsolidation is supposed to occur when the retrieval experience is similar but not identical (Biedenkapp & Rudy, 2004) to the original learning. Yet, a retrieval session that is too different from the original learning procedure might not cause destabilization of the original memory trace (Hunsaker & Kesner, 2013) but instead initiate the formation of a new memory trace such as in extinction learning (Bos et al., 2012). Without an independent index of memory destabilization other than the memory enhancing or amnesic effects of the manipulations themselves, determining the degree of similarity (or dissimilarity) between learning and retrieval presents a problem for empirical falsifiability (Finnie & Nader, 2012).

Criteria for optimal (dis)similarity cannot be inferred from the expression of the target memory itself during memory retrieval, because the mechanisms that mediate memory destabilization are independent from the behavioural fear expression (Mamou, Gamache, & Nader, 2006; Sevenster et al., 2012a). In addition, a certain reactivation procedure may induce plasticity after one but not another learning procedure. We demonstrate that PE can be used as an independent measure of memory destabilization. When there was no modification in CS-US expectancies from acquisition to test, the memory trace was not updated. We believe that, at least in the current protocol, it would be difficult to assess PE driven learning at the moment of reactivation because a small decrease in CS-US expectancies during the memory retrieval session itself would already
induce extinction learning. Then, updating may no longer affect the original memory trace but — because of the small degree of similarity between acquisition and retrieval — instate the formation of a new extinction memory. Because reconsolidation of memory traces corresponding to different response systems (amygdala dependent startle potentiation and hippocampal-dependent declarative memory) calls for different reactivation conditions (Soeter & Kindt, 2010, 2011b), we are now capable of independently assessing the prerequisite for fear memory destabilization in humans, in a non-invasive manner. Conditions that were previously regarded as constraints on reconsolidation (such as too little/too much similarity) may be resolved by taking into account PE during memory retrieval. The assessment of PE provides a feasible tool to develop and optimize reconsolidation-based treatments for patients suffering from chronic relapsing disorders such as anxiety disorders and substance-abuse disorders.
Supplementary Material

MATERIALS AND METHODS

Participants
Forty-five (14 male; 31 female) healthy individuals participated in the study. Mean age was 20.51 years (SD = 1.97), ranging in age between 18 and 26 years. All participants were free from any condition contraindicative to the administration of 40 mg propranolol (S1). Participants received either partial course credit or a small amount of money (€ 35,-) for their participation. All participants gave informed consent and were notified that they could withdraw from participation at any time. The study had full ethical approval. Participants were randomly assigned to one of the three conditions: Negative PE (n = 15, 6 male), No PE (n = 15, 4 male) or Positive PE group (n = 15, 4 male) with the restriction that groups were matched on spider fear, anxiety sensitivity and trait anxiety as was assessed by the Fear of Spiders Questionnaire (FSQ), the Anxiety Sensitivity Index (ASI) and the Trait Anxiety Inventory (STAI-T), respectively.

Apparatus
Stimuli. The conditioned stimuli (CS) consisted of 2 different fear-relevant images depicting a spider and a gun (IAPS, nr 1200; 6210). The startle probe, a 40 ms duration noise burst (104 dB) with a rise/fall time shorter than 1 ms, was presented binaurally through headphones. CS duration was 8 s and the startle probe was delivered 7 s after CS onset, followed by the US 500 ms later. The US consisted of an electrical stimulus (2 ms), determined individually to be ‘uncomfortable though not painful’. Electrical stimulation was delivered through a pair of Ag electrodes of 20 by 25 mm with a fixed inter-electrode mid-distance of 45 mm. Shock deliverance was controlled by a constant current stimulator. Between the electrodes and the skin a conductive gel was applied. Inter-trial intervals (ITI) varied from 15 s to 25 s with an average of 20 s.

Fear potentiated startle. Startle response was measured through electromyography (EMG) of the right orbicularis oculi muscle. Two 5-mm Ag/AgCl electrodes filled with a conductive gel were positioned approximately 1 cm under the pupil and 1 cm below the lateral canthus, respectively; a ground electrode was placed on the forehead, 1 cm below the hairline (Blumenthal et al., 2005). The
EMG signal was sampled at 1000 Hz and amplified in two stages. The input stage had an input resistance of 10 MOhm, a frequency response of DC-1500 Hz and an amplification factor of 200. A 50 Hz notch filter was used to reduce interference of the mains noise. The second stage amplified the signal with a variable amplification factor of 0–100 x and integrated the signal. The raw EMG data were band-pass filtered (28–500 Hz, Butterworth, 4th order (Blumenthal et al., 2005)) to obtain the cleanest possible data without affecting response amplitude. Peak blink amplitude was determined in a 30–120 ms interval following probe onset. Startle response were recorded with the software program VSSRP98.

**Online US-expectancy ratings.** US-expectancy was measured online during each CS presentation, on an 11-point scale ranging from ‘certainly no electric stimulus’ (-5) through ‘uncertain’ (0) to ‘certainly an electric stimulus’ (5). The scale was placed at the bottom of the screen below the CS picture. Participants rated US-expectancy levels by shifting the cursor on the scale with use of the mouse and confirmed their ratings by pushing the left mouse button within 7 s following stimulus onset, before the onset of the startle probe.

**Drug treatment.** Propranolol HCl (40 mg) pills were prepared by a pharmacy. We measured blood pressure with a cuff attached to the right upper arm, using an electronic sphygmomanometer.

**Subjective assessments.** The Fear of Spiders Questionnaire (FSQ) (Szymanski & O’Donohue, 1995) was used to assess the degree of spider fear. In addition, the Anxiety Sensitivity Index (ASI) (Peterson & Reiss, 1992) was taken to assess a subject’s tendency to respond anxiously to the temporary symptoms of the use of propranolol. State and trait anxiety were measured with the State and Trait Anxiety Inventory (STAI-S/STAI-T) (Spielberger et al., 1970) to assess the influence of propranolol on state anxiety and match the groups on general level of anxiety, respectively. Evaluation of the US was assessed on an 11-point scale ranging from ‘unpleasant’ (-5) to ‘pleasant’ (5) to investigate the effect of pill administration on the course of US-evaluation.

**Procedure**
The experiment consisted of three testing sessions on consecutive days (Fig. 1A). Each testing session started with startle habituation trials to stabilize baseline startle reactivity. Startle probes alone (Noise Alone; NA) were presented in addition to the CS presentations to assess baseline startle responding during the experimental phases. Throughout all the experimental phases participants rated their US-
expectancies during each CS presentation. Before testing participants were screened to be free from any medical or psychiatric condition contraindicative to the use of propranolol. Baseline blood pressure and heart rate were measured before participants were subjected to a simple exercise test of cardiovascular function (1 min two-step test). After completion of the two-step test, HR was measured again. Participants (n = 2) whose HR did not increase from pre to post two-step test measurement were excluded from further participation. Finally, participants filled in the FSQ, ASI, and STAI-T questionnaires and gave written informed consent.

**Fear conditioning.** Participants were seated in a chair in front of the computer screen (50 cm distance) and the EMG electrodes were attached. Because we employed an instructed acquisition paradigm in two of the three groups, we assessed preconditioning startle reactivity to the stimuli. Single CS1, CS2 and NA preconditioning trials were preceded by 10 startle habituation trials. The CSs consisted of two different fear-relevant images depicting a spider and a gun (IAPS, nr 1200; 6210). After preconditioning the experimenter entered the room and attached the shock-electrodes to the wrist. Shock intensity was determined individually to be ‘uncomfortable though not painful’. Participants were randomly assigned to any of the three groups: Negative PE, No PE, or Positive PE. During acquisition, in both the Negative PE and No PE group, one of the pictures (CS1) was always paired with the shock (100% of the trials), whereas the other picture was never paired with a shock (CS2) (Fig. 1B). Also, to acquire asymptotic levels of learning, these participants were instructed before the conditioning procedure which picture was followed by the shock (US) and which picture was never followed by the shock. Acquisition consisted of three presentations of each CS and three NA presentations. In the Positive PE group one of the pictures (CS1) was paired with the shock only on the second trial out of three trials (33% of the trials), whereas the other picture was never paired with a shock (CS2) (Fig. 1B). Participants in this group received the instruction that one of the pictures was sometimes followed by the shock and one of the pictures was never followed by the shock. In this group participants had to learn which picture was actually followed by a shock. Before the actual conditioning phase participants were again presented with five startle probes (NA) to reduce general startle reactivity due to exposure to the US during the shock intensity set up procedure. Assignment of the pictures as CS1 or CS2 was counterbalanced across participants. Prior to the fear conditioning phase all participants were instructed to rate their US-
expectancies within the 7 s following each CS onset by moving a cursor on a continuous scale located on the bottom of the screen.

**Memory reactivation.** The memory reactivation session took place 24 h after fear learning. All participants were told that the same pictures would be presented again and they were instructed to remember what they had previously learned (CS-US contingencies). Previous to memory reactivation, 10 startle habituation trials were presented to stabilize baseline startle reactivity. The memory was reactivated with either a single unreinforced (Negative PE) or a single reinforced (No PE and Positive PE) CS1 presentation of 8 s (Fig. 1B). After memory reactivation all participants received (blind) an oral dose of propranolol (40 mg). State anxiety and blood pressure were measured before and 90 min after pill administration.

**Extinction training and reinstatement test.** One day later, on day 3, the procedure was similar for the three groups. Participants were instructed that the same two pictures would be presented again. The instructions did not reveal anything regarding the occurrence of the US. The testing session started again with 10 startle habituation trials to stabilize baseline startle reactivity. During extinction, participants were presented with 12 unreinforced CS1 and CS2 trials and 12 NA trials. After extinction learning, three unsignaled reminder shocks were administered to the wrist, followed by six unreinforced CS1 and CS2 trials and 6 NA trials to test reinstatement of fear. At the completion of the test phase participants evaluated the US.

**Data Analysis**

Startle response outliers were defined by means of within-participants Z scores (\(Z > 3.29\)) (calculated across the three testing days) and replaced by linear trend at point. Startle responses and US-expectancy ratings were subjected to a mixed analysis of variance for repeated measures (ANOVA) with group (Negative PE; No-PE; Positive PE) as between-subjects factor and stimulus (CS1 vs. CS2; CS1 vs. NA) and trial (stimulus presentation) as within-subjects factors. Follow-up analyses were first performed with contrasts as between-subjects factor, such that the groups were similar on acquisition but different on reactivation (Negative PE vs. No PE) or different on acquisition but similar on reactivation (No-PE vs. Positive PE) and stimulus (CS1 vs. CS2; CS1 vs. NA) and trial (stimulus presentation) as within-subjects factors. Second, if appropriate, analyses were performed for the Negative PE,
Positive PE and No PE group separately with stimulus (CS1 vs. CS2; CS1 vs. NA) and trial (stimulus presentation) as within-subjects factors.

US-intensity, FSQ, ASI and STAI-T scores were subjected to ANOVAs with group as between-subjects factor. Mixed ANOVAs with group as between-subjects factor and moment (day 1 vs. day 3; before vs. after pill administration) as within-subjects factor were used to analyse the effect of pill administration on the course of US-evaluation, blood pressure, and STAI-S scores. A Greenhouse-Geisser procedure was used in case of violation of the sphericity assumption in ANOVAs. The alpha level was set at .05 for all statistical analyses.

RESULTS

Questionnaires, Evaluations and Blood Pressure

The groups did not differ in reported US-intensity, US-evaluation, spider fear, anxiety sensitivity or trait anxiety ($F_s < 1.03$) (Table 1). The individually set shock intensity ranged from 4 to 60 mA ($M = 21.84$, $SD = 11.98$). There were no differences in the evaluation of the US between the groups on either day 1 or day 3 ($F_i < 1.36$), indicating that participants experienced the US similarly. In addition, we found no differences between the groups in the change of state anxiety before and after pill administration (moment x group, $F_{i} < 1.18$).

Finally, analysis revealed no differences in the course of systolic and diastolic blood pressure before and after pill intake between the groups (moment x group, $F_i < 1.48$). Both systolic (main effect moment, $F_{i, 42} = 72.10$, $p < .001$, $\eta^2_p = .63$) and diastolic blood pressure (main effect moment; $F_{i, 42} = 12.35$, $p < .001$, $\eta^2_p = .23$) decreased from the first (before pill intake) $(t=0)$ to the second measurement (90 min after pill intake) $(t=1)$ in all groups (Table 2).

On-line US-expectancy Ratings

Acquisition. There was a difference between the three groups in differential responding on the last trial of acquisition (trial 3) (stimulus x group, $F_{i, 42} = 54.21$, $p_i < .001$, $\eta^2_p = .72$). As could be expected from the identical acquisition procedure in the Negative PE and the No PE group, we observed no difference between those groups in differential ratings on the last acquisition trial (trial 3) (stimulus x group, $F_{i, 28} < 1$). In both groups (Negative PE and No PE), we observed acquisition of differential US-expectancy ratings, evidenced by higher ratings to the CS1 compared to the CS2 (main effect stimulus, $F_{i, 28} = 4458.90$, $p < .001$, $\eta^2_p = .99$). While differential ratings on the last trial of acquisition were less pronounced in
Table S1. Mean values (SD) of the US-intensity, US-evaluation, reported spider fear (FSQ), anxiety sensitivity (ASI), and trait anxiety (STAI-T) for the Negative PE (n = 15), No PE (n = 15), and the Positive PE group (n = 15).

<table>
<thead>
<tr>
<th></th>
<th>Negative PE</th>
<th>No PE</th>
<th>Positive PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-intensity (mA)</td>
<td>25.0 (12.7)</td>
<td>21.8 (12.9)</td>
<td>18.7 (10.1)</td>
</tr>
<tr>
<td>US-evaluation</td>
<td>-3.2 (.9)</td>
<td>-3.1 (1.3)</td>
<td>-2.9 (1.1)</td>
</tr>
<tr>
<td>FSQ</td>
<td>32.7 (14.7)</td>
<td>34.8 (27.7)</td>
<td>31.5 (21.2)</td>
</tr>
<tr>
<td>ASI</td>
<td>12.7 (5.3)</td>
<td>11.1 (5.6)</td>
<td>10.0 (5.5)</td>
</tr>
<tr>
<td>STAI-T</td>
<td>39.5 (10.5)</td>
<td>39.2 (7.9)</td>
<td>37.1 (8.0)</td>
</tr>
</tbody>
</table>

the Positive PE group compared to the No PE group (stimulus x group, $F_{(1, 28)} = 60.00, \ p < .001, \ \eta^2_p = .68$), yet the partial reinforcement scheme did result in significant differential ratings in the Positive PE group (main effect stimulus, $F_{(1, 14)} = 104.40, \ p < .001, \ \eta^2_p = .88$).

**Memory reactivation.** Analysis revealed a difference between the three groups on US-expectancy ratings from the end of acquisition (trial 3) to the reactivation trial (trial x group, $F_{(1, 42)} = 8.46, \ p < .001, \ \eta^2_p = .29$). While the Negative PE and No PE group did not differ (trial x group, $F_{(1, 28)} < 1$; main effect trial, $F_{(1, 28)} < 2.70$), we did observe a difference between the Positive PE and No PE group on US-expectancy ratings from the end of acquisition (trial 3) to the reactivation trial (trial x group, $F_{(1, 28)} = 13.30, \ p < .001, \ \eta^2_p = .32$). Ratings to the CS1 remained similar in the No PE group (main effect trial, $F_{(1, 14)} < 1.24$), while there was an increase from the end of acquisition (trial 3) to the reactivation trial in the Positive PE group (main effect trial, $F_{(1, 14)} = 23.24, \ p < .001, \ \eta^2_p = .62$).

**Extinction and reinstatement.** Analysis revealed a difference between the three groups in differential US-expectancy ratings (CS1 vs. CS2) during extinction learning (trial 1 vs. trial 12) (stimulus x trial x group, $F_{(1, 42)} = 4.66, \ p < .015, \ \eta^2_p = .18$). Extinction learning of US-expectancy ratings was either facilitated (Negative PE vs. No PE; stimulus x group, $F_{(1, 28)} = 10.12, \ p < .004, \ \eta^2_p = .27$) or impeded (Positive PE vs. No PE; stimulus x trial x group, $F_{(1, 28)} = 11.29, \ p < .002, \ \eta^2_p = .29$) relative to the group in which updating did not occur (No PE). The three groups differed in differential US-expectancy ratings on the first trial of extinction (stimulus x group, $F_{(1, 42)} = 5.56, \ p$
Table S2. Mean values (SD) of the systolic blood pressure (BP), diastolic blood pressure and state anxiety before (t = 0) and 90 min after (t = 1) pill administration for the Negative PE (n = 15), No PE (n = 15), and the Positive PE group (n = 15).

<table>
<thead>
<tr>
<th></th>
<th>Negative PE</th>
<th></th>
<th>No PE</th>
<th></th>
<th>Positive PE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t =0</td>
<td>t=1</td>
<td>t=0</td>
<td>t=1</td>
<td>t=0</td>
</tr>
<tr>
<td>Systolic BP</td>
<td>115.0</td>
<td>(15.0)</td>
<td>112.3</td>
<td>101.3</td>
<td>118.2</td>
</tr>
<tr>
<td>Diastolic BP</td>
<td>69.3</td>
<td>64.0</td>
<td>68.3</td>
<td>66.9</td>
<td>73.9</td>
</tr>
<tr>
<td>STAI-S</td>
<td>33.4</td>
<td>31.1</td>
<td>34.8</td>
<td>31.1</td>
<td>31.1</td>
</tr>
<tr>
<td></td>
<td>(6.3)</td>
<td>(5.9)</td>
<td>(7.9)</td>
<td>(5.5)</td>
<td>(6.1)</td>
</tr>
</tbody>
</table>

< .007, $\eta^2_p = .21$. We observed greater differential ratings on the first trial of extinction in the No PE group relative to the Negative PE (stimulus x group, $F_{(1, 28)} = 8.99, p < .006, \eta^2_p = .24$), while the difference between the No PE group and Positive PE group approached significance (stimulus x group, $F_{(1, 28)} = 3.24, p < .082, \eta^2_p = .10$), respectively. In spite of these differences, memory for CS-US contingencies was still intact in all the groups on the first trial of extinction (day 3), as was evidenced by more pronounced US-expectancy ratings to the CS1 compared to the CS2 at the beginning of extinction (trial 1) in all three groups: the Negative PE group (main effect stimulus, $F_{(1, 14)} = 80.94, p < .001, \eta^2_p = .85$), the No PE group (main effect stimulus, $F_{(1, 14)} = 417.00, p < .001, \eta^2_p = .97$) and the Positive PE group (main effect stimulus, $F_{(1, 14)} = 396.00, p < .001, \eta^2_p = .97$). There was also a difference between the three groups in reinstatement of US-expectancy ratings from the end of extinction (trial 12) to the test trial (stimulus x trial x group, $F_{(1, 42)} = 3.31, p < .046, \eta^2_p = .14$). The unpredicted shocks resulted in an increase from the end of extinction (trial 12) to the first test trial in both the Negative PE and the No PE group (stimulus x trial x group, $F_{(1, 28)} < 1$; stimulus x trial, $F_{(1, 28)} = 27.40, p < .001, \eta^2_p = .50$). However, while we observed a difference in the predicted direction between the No PE and Positive PE group in reinstatement of US-expectancy ratings (stimulus x trial x group, $F_{(1, 28)} = 3.45, p < .074, \eta^2_p = .11$), a resistance to extinction observed in the Positive PE group prevented the proper assessment of reinstatement of US-expectancy ratings in this group (stimulus x trial, $F_{(1, 14)} < 1.97$).
Fear Potentiated Startle

Habituation. The groups did not differ in startle responding over the course of habituation during the habituation sessions on day 1, day 2 and day 3 (trials 1 to 10; trials 1 to 5; trial x group, \( F_\text{S} < 1.49 \)).

Noise Alone trials The three groups differed in startle responding to the NA trials during acquisition (trials 1 to 3) (trial x group, \( F_\text{S} = 3.12, p < .019, \eta^2_p = .13 \)). The Negative PE and No PE group did not differ (trials 1 to 3, trial x group: \( F_{12., 56} < 1.73 \)), while we did observed a difference in startle response to the NA trials during acquisition between the Positive PE group compared to the No PE group (trial x group, \( F_{12., 56} = 4.66, p < .013, \eta^2_p = .14 \)). Thus, uncertainty with respect to CS-US contingencies augmented initial NA startle responding in the group in which acquisition was partially reinforced (Positive PE). The groups did not differ in startle response to the NA trials during extinction (trials 1 to 12) and reinstatement (trials 1 to 6) (trial x group, \( F_\text{S} < 1.07 \)).

Preconditioning. Given that we employed an instructed acquisition paradigm, we assessed preconditioning startle reactivity to the CS1, CS2 and NA. Differential responding (CS1 vs. CS2) differed between the three groups during preconditioning (stimulus x group, \( F_{2, 42} = 4.67, p < .015, \eta^2_p = .18 \)). The Negative PE and No PE group did not differ in differential responding during preconditioning (stimulus x group, \( F_{11, 28} < 1 \)), while there was a difference between the No PE and Positive PE group (stimulus x group, \( F_{11, 28} = 10.44, p < .003, \eta^2_p = .27 \)). The No PE group showed higher responding to the CS1 compared to the CS2 (main effect stimulus, \( F_{11, 14} = 10.90, p < .005, \eta^2_p = .44 \)), whereas we observed a trend towards higher responding to the CS2 compared to the CS1 in the Positive PE group (main effect stimulus, \( F_{11, 14} = 2.97, p < .11, \eta^2_p = .18 \)).

Fear conditioning. In spite of the initial difference in startle responding on the preconditioning trial, the groups did not differ in startle responding (CS1 vs. CS2) on the first conditioning trial (stimulus x group, \( F_{2, 42} < 1 \); main effect stimulus, \( F_{1, 42} < 1 \)). Therefore, acquisition of fear conditioning was assessed from the first acquisition trial to the last (trial 3). There was indeed successful fear conditioning on day 1, as indicated by an increase in differential startle responding (CS1 vs. CS2) from the first to the last trial of acquisition (stimulus x trial, \( F_{1, 42} = 10.01, p < .003, \eta^2_p = .19 \)), that did not differ between the groups (stimulus x group, \( F_{2, 42} < 1 \)). On the last trial of acquisition there was more startle potentiation to the CS1 compared to the CS2 (main effect stimulus, \( F_{1, 42} = 23.04, p < .001, \eta^2_p = .35 \)) which did not differ between the three groups (stimulus x group, \( F_{2, 42} < 1 \)).
Memory reactivation. Analysis of the startle fear response on day 2 revealed stronger responding to the CS1 compared to the NA trial (main effect stimulus, $F(1, 42) = 11.31, p < .002, \eta^2_p = .21$), which did not differ between the groups (stimulus x group, $F(1, 42) < 2.04$). Thus, startle responding was intact irrespective of whether PE and subsequent memory destabilization occurred. This is in line with previous studies in which fear memory expression during reactivation is dissociated from the mechanisms that allow for the updating of memory (Mamou et al., 2006; Sevenster et al., 2012a).

Extinction. Because differential fear responding was eliminated at the first retention trial (first trial of extinction), no extinction learning (trial 1 vs. 12) of differential startle responding (CS1 vs. CS2) was observed either in the Negative PE (stimulus x trial x group, $F(1, 28) = 6.86, p < .014, \eta^2_p = .20$) or in the Positive PE group (stimulus x trial x group, $F(1, 28) = 6.34, p < .018, \eta^2_p = .19$) compared to the No PE group. Extinction learning approached significance in the No PE group, evidenced by a decrease in differential responding (stimulus x trial, $F(1, 14) = 4.04, p < .064, \eta^2_p = .22$), while no extinction learning was observed either within the Negative PE (stimulus x trial, $F(1, 14) < 3.11$) or within the Positive PE group (stimulus x trial, $F(1, 14) < 2.36$). By the end of the extinction phase (trial 12) differential startle fear responding was absent in all groups (main effect stimulus, $F(1, 42) < 2.20$; stimulus x group, $F(2, 42) < 1$).