Fear memory uncovered: Prediction error as the key to memory plasticity

Sevenster, D.

Link to publication

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

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3

Instructed extinction differentially affects the emotional and cognitive expression of associative fear memory

Dieuwke Sevenster
Tom Beckers
Merel Kindt

Psychophysiology, 2012, 49: 1426-1435
ABSTRACT – Instructed extinction after fear conditioning is relatively effective in attenuating electrodermal responding. Testing the single process account of fear learning, we examined whether this manipulation similarly affects the startle response. Skin conductance responses, startle responses and online US-expectancy ratings were measured during fear acquisition (day 1), extinction and reinstatement (day 2). Before extinction onset, half of the subjects were instructed that the CS would not be followed by the US (Instructed Extinction), while the other subjects were not instructed (Normal Extinction). This simple instruction completely abolished both differential SCR and US-expectancy ratings, but not the startle fear response. While the manipulation facilitated extinction learning, it did not prevent the recovery of the startle response. The present findings are better explained by a dual rather than a single process account of fear learning.
BACKGROUND

Fear learning prepares an individual for action when confronted with danger. Once a fear memory is established it appears relatively resistant to change. From an evolutionary perspective this is highly adaptive as it allows for rapid responding when the individual faces the feared object. The human fear conditioning paradigm is well suited to study the formation of associative fear memory and its subsequent expression. In a prototypical fear conditioning procedure, a previously neutral or ambiguous stimulus (Conditioned Stimulus; CS) acquires the capacity to elicit a robust conditioned response (CR; e.g., startle potentiation or electrodermal responding) after being paired with an intrinsically aversive consequence (Unconditioned Stimulus; US), such as an electric shock (Pavlov, 1927). The standard procedure to reduce the conditioned fear response is extinction: the repeated presentation of the feared stimulus without its adverse consequence (CS-noUS).

Several studies suggest that the explicit instruction that the CS will no longer be followed by the US results in an immediate elimination of conditioned responding—even stronger or more promptly than by experience (Hugdahl & Öhman, 1977; Hugdahl, 1978; Lipp & Edwards, 2002; cf. Lovibond & Shanks, 2002). The finding that such a simple cognitive manipulation affects conditioned responding supports a single process account of fear learning. According to this account both affective and expectancy learning originate from a single learning system (Brewer, 1974; Dawson & Schell, 1985; Lipp & Purkis, 2005; Lovibond & Shanks, 2002; Lovibond, 2004; Mitchell et al., 2009). Affective learning refers to the mechanism by which affective valence from the US is transferred to the CS; during expectancy learning the CS gains the capacity to predict the occurrence of the US (Baeyens et al., 1995; Purkis & Lipp, 2001). According to the single process account, affective learning can easily be modified by verbal instructions (Mitchell et al., 2009). This contrasts with the dual process account of fear learning, which holds that affective learning is mediated by a learning mechanism that is distinct from the one that underlies expectancy learning (Baeyens et al., 1995; Hamm & Vaitl, 1996; Hamm & Weike, 2005). According to the dual process account, affective and expectancy learning can occur separately from one another.

Findings on instructed extinction point towards a single underlying mechanism as changes in contingencies lead to the immediate reversal of conditioned responses. However, instructional effects on conditioned fear responding have mainly been investigated by use of the skin conductance
response (SCR). It has been argued that SCR, although traditionally regarded as a measure of fear (Prokasy & Kumpfer, 1973), cannot be considered as a specific index for fear learning (Hamm & Vaitl, 1996; Hamm & Weike, 2005; Weike et al., 2007). In contrast, the startle response, an automatic defensive reflex, corresponds to the affective component of fear learning (Davis, 2006; LeDoux, 2000). Indeed, several studies show that multiple response systems seem to be involved in the formation and expression of associative fear memory and that the physiological correlates of these response systems can dissociate (Hamm & Weike, 2005; Knight, Smith, Cheng, Stein, & Helmstetter, 2004; Sevenster et al., 2012a; Soeter & Kindt, 2010, 2011b, 2012; Squire, 2004). Thus, while electrodermal responding is susceptible to verbal instructions, similar results should not necessarily be observed with respect to the startle response. Yet, from the single process account of fear learning it follows that the startle response should be equally susceptible to instructions regarding the CS-US contingencies.

In the present experiment we investigated the sensitivity of both SCR and startle response to instructed extinction. A human differential fear conditioning paradigm was used in which pictures of geometrical stimuli served as CSs. We presented neutral stimuli because startle responses to fear-relevant stimuli are known to be resistant to extinction (Öhman, Erixon, & Lofberg, 1975). Fear learning was established on day 1 by pairing one of the CSs with an aversive electrical shock while startle response, SCR and online US-expectancy were measured. One day later, half of the participants (Instructed Extinction Group) were instructed that the pictures would no longer be followed by the shock while the other participants (Normal Extinction Group) did not receive instructions. Different from earlier studies on instructed extinction (Hugdahl & Öhman, 1977; Hugdahl, 1978; Lipp & Edwards, 2002), the US-electrodes remained attached to the participant’s wrist during the extinction phase. First, this allowed for a more pure assessment of instructional effects as it excludes a possible generalization decrement of conditioned responding due to context changes. Second, leaving the electrodes attached allowed us to investigate the effect of verbal instructions, not only on extinction learning, but also on the return of fear after presenting the participants with unsignaled shocks (reinstatement testing). The single process account of affective learning predicts that the extinction instructions will affect the emotional expression of associative fear learning as readily as SCR, indicated by a direct elimination of differential startle fear response. Also, if the instructions prevent the recovery of both US-expectancy ratings and SCR after reinstatement shocks, the
startle response will follow this pattern, such that reinstatement of differential startle response will not be observed.

MATERIALS AND METHODS

Participants
Forty (9 male; 31 female) healthy undergraduate students participated in the study, ranging in age between 18 and 28 years (M = 21.75, SD = 2.60). Participants received either partial course credit or a small amount of money for their participation. All participants gave informed consent and were notified that they could withdraw from participation at any time. The Ethics Committee of the University of Amsterdam approved the study. Participants were randomly assigned to either the Instructed Extinction group (n = 20; 4 males) or the Normal Extinction group (n = 20; 5 males).

Apparatus

Stimuli. The conditioned stimuli (CSs) consisted of 2 different images depicting geometrical figures of a blue square and a yellow circle. 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 Digitimer DS7A constant current stimulator (Hertfordshire, UK). Between the electrodes and the skin a conductive gel (Signa, Parker) was applied.

Fear potentiated startle (FPS). Startle response was measured through electromyography (EMG) of the right orbicularis oculi muscle. Two 5 mm Ag/AgCl electrodes filled with a conductive gel (Signa, Parker) 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 hairline (Blumenthal et al., 2005). The startle probe was a 40 ms duration noise burst (104 dB) with a rise/fall time shorter than 1 ms, which was presented binaurally through headphones (Sennheiser, model HD 25-1 II). 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 30–150 ms interval following probe onset.

**Skin conductance response (SCR).** Electrodermal activity was measured using an input device with a sine-shaped excitation voltage (1 V peak-peak) of 50 Hz, which was derived from the mains frequency. Two Ag/AgCl electrodes of 20 by 16 mm were attached with adhesive tape to the medial phalanges of the first and third fingers of the non-preferred hand. The SCR signal was sampled at 1000 Hz. The signal from the input device was led through a signal-conditioning amplifier and the analogue output was digitized at 100 Hz by a 16-bit AD-converter (National Instruments, NI-6224). Startle response and electrodermal activity were recorded with the software program VSSRP98. Phasic electrodermal responding to the CS was calculated by subtracting the baseline (mean skin conductance level during the 2 s period before stimulus onset) from the maximum score, determined at 0.5 s intervals, during the 0 to 7 s window after CS onset (entire-interval response, EIR) (Pineles et al., 2009). This is a well-established approach of examining electrodermal reactivity and has been used extensively in human psychophysiological research (Milad et al., 2005; Orr et al., 2000; Pineles et al., 2009; Raes et al., 2011).

**Online US-expectancy ratings.** US-expectancy was measured online during each image 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.

**Subjective assessments.** Evaluation of the US was assessed on an 11-point scale ranging from “unpleasant” (-5) to “pleasant” (5). General level of anxiety was measured with the Trait Anxiety Inventory (STAI-T; Spielberger, Gorsuch, & Lushene, 1970).

**Procedure**
After giving informed consent participants were seated in front of a computer screen (50 cm distance) in a sound attenuated room. The EMG, SCR and shock electrodes were attached and US-intensity level was determined by gradually increasing shock intensity (starting at 1 mA) until subjects indicated the shock to be ‘uncomfortable though not painful’. The experiment consisted of two testing
sessions on consecutive days. Each testing session started with ten startle habituation trials to stabilize baseline startle reactivity. To assess baseline startle responding during the experimental phases startle probes alone (Noise Alone; NA) were presented in addition to the CS presentations. Throughout all the experimental phases participants rated their US-expectancy during each CS presentation. For a schematic representation of the experimental design see Figure 1A.

**Fear conditioning.** All participants underwent fear conditioning on day 1. The CSs consisted of 2 different images depicting geometrical figures. One of the images (CS1) was paired with a mild shock to the wrist (US) on four out of six trials, whereas the other picture was never paired with a shock (CS2). Assignment of the pictures as CS1 or CS2 was counterbalanced across participants. Both CSs were presented 6 times for 8 s. A startle probe (40 ms; 104 dB) was delivered 7 s after CS onset, followed by the US after another 500 ms. The US consisted of an electrical stimulus (2 ms). Note that delivery of neither the startle probe nor the US interfered with measurement of SCR as maximum SCR score was determined during 7 s following stimulus onset. For a schematic representation of a conditioning trial see Figure 1B. Intertrial intervals (ITI) varied from 15 s to 25 s with an average of 20 s. All participants were instructed that one of the pictures was followed by a shock on most trials, while the other picture was never followed by a shock. It was stressed that the latter was by no means reinforced on the next day. Also, it was pointed out that they should learn the CS-US contingency on basis of the images. Participants rated their US-expectancy during each image presentation by moving the bar on the scale to the corresponding value ranging from ‘certainly no electric stimulus’ (−5) through ‘uncertain’ (0) to ‘certainly an electric stimulus’ (5).

**Extinction training and reinstatement test.** On day 2, participants were randomly assigned to either the Instructed Extinction group or the Normal Extinction group. After attachment of the SCR, EMG and shock electrodes, all participants were told that the same two pictures of geometrical figures would be presented again. Crucially, participants in the Instructed Extinction group were explicitly instructed that the pictures would not be followed by shock. Participants in both groups were presented with 16 unreinforced CS1 and CS2 trials. After extinction learning, three unsignaled reminder shocks were administered to the wrist, followed by 8 unreinforced CS1 and CS2 trials to test reinstatement of fear. Again, participants rated their US-expectancy online. Finally, subjects filled in the trait anxiety (STAI-T) questionnaire and rated US-pleasantness.
Figure 1. (A) Schematic representation of the experimental design. (B) A conditioning trial of a reinforced stimulus presentation. Stimulus duration was 8 s. A startle probe was delivered 7 s after CS onset, followed by the US (2 ms) after 500 ms. Participants rated online US expectancies during the 7 s following CS onset. Peak skin conductance response was calculated for the 0–7 s window following CS onset.

Data Analysis
STAI-T scores, US-intensity and US-evaluation scores were subjected to ANOVAs with group (Normal Extinction vs. Instructed Extinction) as between-subjects factor. Fear potentiated startle and skin conductance responses were standardized using within-participants Z-scores (calculated across the two testing days). US-expectancy ratings, startle responses and electrodertal activity were then subjected to a mixed analysis of variance for repeated measures (ANOVA) with group (Instructed Extinction vs. Normal Extinction) as between-subjects factor and stimulus (CS1 vs. CS2) and time (stimulus presentation) as within-subjects factors. The alpha level was set at .05 for statistical analyses. Planned comparisons were performed for each group separately. A Greenhouse-Geisser procedure was used.
in case of violation of the sphericity assumption in ANOVAs. To assess the speed of extinction learning ANOVAs were performed. We averaged the data over blocks of two trials to reduce the variability of physiological responding. We then performed multiple comparisons (CS1 vs. CS2) for each separate block (1 to 8) of extinction. The p value was corrected using the Benjamini–Hochberg false discovery technique (false detection rate, FDR) for multiple test comparisons (Benjamini & Hochberg, 1995).

RESULTS

Questionnaires and Evaluations

The groups did not differ in trait anxiety ($F_{(1, 38)} < 1.94$). The individually determined shock intensity ranged from 6 to 46 mA ($M = 17.88, SD = 10.88$). Comparing the groups with respect to US-intensity revealed no differences in US intensities between the Instructed Extinction group ($M = 18.65, SD = 9.78$) and the Normal Extinction group ($M = 17.10, SD = 12.07; F_{(1, 38)} < 1$). Additionally, there was no difference in the evaluation of the US on either day 1 or day 2 ($F_{s} < 1.82$), indicating that participants experienced the US similarly.

Manipulation Check

US-expectancy ratings. We observed acquisition of US-expectancy, as evidenced by a significant increase in differential US-expectancy ratings (CS1 vs. CS2) on day 1 (trials 1 to 6; stimulus x trial; $F_{(5, 190)} = 70.07, p < .001, \eta^2_p = .65$; Fig. 2A,B). This pattern did not differ between groups (stimulus x trial x group; $F_{(5, 190)} < 1$). The instruction that the pictures would not be followed by the shock anymore (Instructed Extinction group) reduced the differential US-expectancy ratings from the last trial of acquisition to the first trial of extinction, relative to no instruction (Normal Extinction group; stimulus x trial x group; $F_{(1, 38)} = 89.43, p < .001, \eta^2_p = .70$). Although differential responding somewhat decreased from the last acquisition trial to the first extinction trial (stimulus x trial; $F_{(1, 19)} = 3.86, p = .064, \eta^2_p = .17$), US-expectancy ratings to the CS1 remained significantly higher compared to ratings to the CS2 during the first extinction trial in the Normal Extinction group (main effect of stimulus; $F_{(1, 19)} = 554.45, p < .001, \eta^2_p = .97$; Fig. 2A). In contrast, there was a significant decrease in differential ratings from the last trial of acquisition to the first extinction trial in the Instructed Extinction group (stimulus x trial; $F_{(1, 19)} = 136.91, p < .001, \eta^2_p = .88$). The instruction completely eliminated differential US-expectancy ratings on the first trial of extinction (main effect of stimulus; $F_{(1, 19)} < 1.58$; Fig. 2B).
The course of extinction learning (trials 1 to 16) differed accordingly between the groups (stimulus x trial x group; F(15, 570) = 26.44, p < .001, \(\eta^2_p = .41\)). A differential decrease in US-expectancy ratings was observed during extinction learning in the Normal Extinction group (trials 1 to 16; stimulus x trial; F(15, 285) = 30.79, p < .001, \(\eta^2_p = .62\)). Given that differential US-expectancy was already absent at the beginning of extinction in the Instructed Extinction group, there was no differential change in US-expectancy ratings (trials 1 to 16; stimulus x trial; F(15, 285) < 1).

Figure 2. Mean online US expectancy ratings to the CS1 and CS2 during acquisition, extinction, and reinstatement test for (A) the normal extinction group (n = 20) and (B) the instructed extinction group (n = 20). Error bars represent SEM.
Finally, we observed a difference between groups in differential US-expectancy ratings from the last trial of extinction to the reinstatement test trial (stimulus x trial x group; $F_{(1, 38)} = 7.15$, $p = .011$, $\eta^2_p = .16$). The three reminder shocks reinstated differential US-expectancy ratings in the Normal Extinction group (stimulus x trial; $F_{(1, 19)} = 9.81$, $p = .005$, $\eta^2_p = .34$; Fig. 2A) whereas the instructions prevented the return of differential ratings in the Instructed Extinction group (stimulus x trial; $F_{(1, 19)} < 1.71$; Fig. 2B). These results show that the manipulation was successful, given that differential US-expectancy ratings were immediately absent after the instruction and did not return after the unpredicted shocks in the Instructed Extinction group.

**Fear Potentiated Startle**

An increase in differential (CS1 vs. CS2) startle responding indicated successful acquisition on day 1 (trials 1 to 6; stimulus x trial; $F_{(5, 190)} = 3.56$, $p = .004$, $\eta^2_p = .09$). This pattern did not differ between groups (stimulus x trial x group; $F_{(5, 190)} < 1.09$; Fig. 3A,B). Differential responding persisted from the last trial of acquisition to the first trial of extinction (stimulus x trial; $F_{(1,38)} < 1$). Instructed extinction did not affect the startle response, given that startle potentiation to the first test trials did not decline in either group (stimulus x trial x group; $F_{(1,38)} < 1$). Indeed, on the first trials of extinction (trials 1-2) a significant difference between startle response to the CS1 and the CS2 was still observed in both the Normal Extinction group (main effect of stimulus; $F_{(1, 19)} = 16.46$, $p = .001$, $\eta^2_p = .46$) and the Instructed Extinction group (main effect of stimulus; $F_{(1, 19)} = 8.59$, $p = .009$, $\eta^2_p = .31$). However, the groups differed in overall differential responding during extinction training (stimulus x group; $F_{(1, 28)} = 14.04$; $p = .001$, $\eta^2_p = .27$). Planned comparisons showed a near-significant reduction in differential responding during extinction learning (trials 1 to 16) in the Normal Extinction group (stimulus x trial; $F_{(15, 285)} = 1.67$, $p = .056$, $\eta^2_p = .08$) and a significant decrease in differential responding in the Instructed Extinction group (stimulus x trial; $F_{(15, 285)} = 1.90$, $p = .023$, $\eta^2_p = .09$). Furthermore, the groups differed in speed of extinction learning, as demonstrated by a series of ANOVAs ($p = 0.05$ corrected using false discovery rate, FDR; Benjamini & Hochberg, 1995). In the Normal Extinction group, responding to the CS1 compared to the CS2 remained significant during trials 3-4 (main effect of stimulus; $F_{(1,19)} = 17.07$, $p = .001$, $\eta^2_p = .47$), trials 5-6 (main effect of stimulus; $F_{(1,19)} = 9.04$, $p = .007$, $\eta^2_p = .32$), trials 7-8 (main effect of stimulus; $F_{(1,19)} = 9.52$, $p = .006$, $\eta^2_p = .33$) and trials 9-10 (main effect of stimulus; $F_{(1,19)} = 11.84$, $p = .003$, $\eta^2_p = .38$). Differential responding was only absent
by trials 11-12 and remained absent over the last four trials (trials 13-14 and trials 15-16; main effect of stimulus; $F_{(1, 19)} < 3.43$; Fig. 3A). In contrast, the significant difference between the CS1 and CS2 at the beginning of extinction (trials 1-2) disappeared much faster and was already absent at extinction trials 3-4 (main effect of stimulus; $F_{(1, 19)} < 1$) in the Instructed Extinction group. This pattern persisted throughout the remainder of the extinction procedure (trials 5-6, 7-8, 9-10, 11-12, 13-14, 15-16; main effect of stimulus; $F_{(1, 19)} < 2.34$; Fig. 3B).

**Figure 3.** Mean startle response to the CS1, CS2 and NA during acquisition, extinction, and reinstatement test for (A) the normal extinction group ($n = 20$) and (B) the instructed extinction group ($n = 20$). Error bars represent SEM.
Thus, we observed faster and more robust extinction learning in the Instructed Extinction group than in the Normal Extinction group. While the extinction instruction did not prevent initial startle potentiation, it did increase the speed of extinction learning of the conditioned startle response. Moreover, the three reminder shocks upon completion of extinction yielded a reinstatement of differential startle fear response from the last trial of extinction to the test trial (stimulus x trial; $F_{(1, 38)} = 4.36; p = .044, \eta^2_p = .10$). Although visual inspection suggests that the reinstatement effect was greater in the Normal Extinction group compared to the Instructed Extinction group, we observed no differences in reinstatement between the two groups (stimulus x trial x group; $F_{(1, 38)} < 1$; Fig. 3A,B). In sum, the instructions neither abolished startle potentiation nor prevented the recovery of fear after a reinstatement procedure, even though they facilitated extinction learning.

**Skin Conductance Response (SCR)**

Analysis showed acquisition of electrodermal responding on day 1, as indicated by an increase in differential SCR (trials 1 to 6; stimulus x trial; $F_{(5, 190)} = 6.24; p < .001, \eta^2_p = .14$). Acquisition of SCR did not differ between groups (stimulus x trial x group; $F_{(5, 190)} < 1$; Fig. 4A,B). We observed a near significant difference in the decrease of differential SCR from the last acquisition trial to the first extinction trial between the Normal Extinction and Instructed Extinction group (stimulus x trial x group; $F_{(1, 38)} = 3.41, p = .073, \eta^2_p = .08$). Indeed, differential SCR remained stable when tested 24 hr later in the Normal Extinction group (stimulus x trial; $F_{(1, 19)} = .09, p = .763$) while a near significant decrease was observed in the Instructed Extinction group (stimulus x trial; $F_{(1, 19)} = 4.03; p = .059, \eta^2_p = .08$). Post-hoc comparisons revealed that SCR to the CS1 differed from responding to the CS2 at the first trial of extinction in the Normal Extinction group (main effect of stimulus; $F_{(1, 19)} = 18.46; p < .001, \eta^2_p = .49$; Fig. 4A), while differential SCR was no longer present in the Instructed Extinction group (main effect of stimulus; $F_{(1, 19)} < 1$; Fig. 4B). In addition, overall differential SCR during extinction learning (trials 1 to 16) on day 2 differed between the groups (stimulus x group; $F_{(1, 38)} = 6.05; p = .019, \eta^2_p = .14$). Similar to the US-expectancy ratings, we observed no change in differential SCR in the Instructed Extinction group (stimulus x trial; $F_{(15, 265)} < 1$) and no differential SCR was observed throughout the whole extinction procedure (main effect of stimulus; $F_{(1, 19)} < 1$). However, there was no reduction in differential SCR in the Normal Extinction group either when analysing all extinction trials (stimulus x trial; $F_{(15, 265)} < 1.28$). However, analysis of the
first vs. the last extinction trial revealed that the unreinforced trials actually did extinguish the differential SCR (stimulus x trial; $F_{(1, 19)} = 4.48$, $p = .048$, $\eta^2_p = .19$).

The reminder shocks did not result in reinstatement of differential electrodermal responding from the last trial of extinction to the first test trial (stimulus x trial; $F_{(1, 38)} < 1$). We also observed no difference in responses to the CS1 compared to responses to the CS2 at the first test trial (main effect of stimulus; $F_{(1, 38)} < 1$). Thus, similar to US-expectancies, SCR was not reinstated in the Instructed Extinction group.

![Figure 4](image)

**Figure 4.** Mean skin conductance response (SCR) to the CS1 and CS2 during acquisition, extinction, and reinstatement test for (A) the normal extinction group ($n = 20$) and (B) the instructed extinction group ($n = 20$). Error bars represent SEM.
Extinction group. Unlike the US-expectancies, SCR was also not reinstated in the Normal Extinction group.

**DISCUSSION**

The aim of the present study was to investigate the sensitivity of both SCR and startle potentiation to instructed extinction. The online US-expectancy ratings show that the intended manipulation was successful. Contrary to no instruction (Normal Extinction group), the explicit instruction immediately eliminated the differential US-expectancy, which did not reinstate after the reminder shocks (Instructed Extinction group). In accordance with previous studies utilizing neutral stimuli in a fear conditioning procedure (Hugdahl & Öhman, 1977; Hugdahl, 1978; Lipp & Edwards, 2002), the instructions completely abolished differential electrodermal responding. In contrast, the startle potentiation was less susceptible to verbal instructions: it was still present at the beginning of extinction and the instructions did not prevent the return of fear after the unexpected reinstatement shocks.

Although verbal instructions had an immediate impact on SCR, non-differential electrodermal responding remained high on the first extinction trial. Electrodermal activity on the first extinction trial might reflect an orienting response to the first stimulus presentations after acquisition training. Some studies even disregard the first (Schiller et al., 2009) or the first three test trials (Phelps, Delgado, Nearing, & LeDoux, 2004) in their analyses because of this orienting response. While reinstatement of the startle response was present, regardless of whether participants received verbal instructions, electrodermal responding did not recover after the reminder shocks in the Instructed Extinction group. Neither could we observe the return of electrodermal responding in the Normal Extinction group. Note that this lack of reinstatement could be attributed to the amount of extinction trials. Even though we recently did observe reinstatement of SCR to fear-relevant stimuli after 16 extinction trials (Sevenster et al., 2012a), SCR to neutral stimuli habituates after a certain number of trials. Indeed, studies focusing on only SCR did employ extinction training of 4 to 6 trials to prevent SCR from habituation and allow for the return of electrodermal responding at test (Vervliet, Vansteenwegen, Baeyens, Hermans, & Eelen, 2005).

Evidence for the effectiveness of verbal instructions on extinction of conditioned responding to fear-relevant stimuli is mixed. So far, two studies found evidence for resistance to instructed extinction of SCR for fear-relevant stimuli (Hugdahl & Öhman, 1977; Hugdahl, 1978). To date, however, attempts to replicate
this finding failed to demonstrate that SCR can dissociate from US-expectancies in an instructed extinction paradigm (Eelen, as cited in Cook, Hodes, & Lang, 1986; Dawson, Schell, & Banis, 1986; Lipp & Edwards, 2002). In a recent study, we could also not demonstrate that fear-relevance impedes instructed extinction effects, as SCR to a fear-relevant stimulus was, similar to the retrospective US-expectancy ratings, immediately reduced after verbal instructions, whereas startle responding remained intact (Sevenster et al., 2012a). Even though electrodermal responses to fear-relevant stimuli may exhibit greater resistance to extinction (Dawson et al., 1986; Öhman et al., 1975; but see McNally, 1987), a similar resistance has been observed for US-expectancy ratings (Dawson et al., 1986). Thus, not only for neutral stimuli but also for fear-relevant stimuli the electrodermal conditioning and US-expectancies seem to converge (Dawson et al., 1986; McNally, 1987).

The current results cannot easily be explained by the single process hypothesis. Instead of mirroring the US-expectancy ratings and electrodermal activity, the startle response was differentially affected by the extinction instructions. A similar dissociation between response systems has been observed previously. Conditioning experiments showed that SCR is associated with non-specific increases in arousal. First, electrodermal conditioning occurred when the US consisted of either aversive electrical stimulation, or non-aversive vibrotactile stimulation (Hamm & Vaitl, 1996) or a reaction time task (Lipp, Sheridan, & Siddle, 1994). Thus, the affective valence of the US does not have a modulatory role in acquisition of SCR. Second, electrodermal conditioning requires awareness of the relationship between the CS and its consequence, as only those participants who could correctly recall the contingencies acquired SCR conditioning (Hamm & Vaitl, 1996). SCR reflects an increase in orienting activity, triggered in response to a significant event—regardless of whether it is positive or negative. Skin conductance conditioning is, thus, supposed to be an indicator of anticipatory arousal and is closely associated with declarative knowledge. This contrasts sharply with the startle response that is considered to be a reliable measure of fear learning (Hamm & Weike, 2005; LeDoux, 2000). Potentiation of the startle reflex was only observed when the US consisted of aversive electrical stimulation, as opposed to non-aversive vibrotactile stimulation (Hamm & Vaitl, 1996) or a reaction time task (Lipp et al., 1994). Furthermore, the startle response seems to be automatically potentiated by the CS, since startle potentiation can occur in absence of CS-US contingency awareness (Hamm & Vaitl, 1996). This apparent dissociation between response systems suggests that, contrary to SCR, the startle response is an index of
aversive conditioning. In further support of diverging response systems, our previous studies on human fear memory reconsolidation demonstrated a dissociation between the startle response on the one hand and US-expectancies and SCR on the other hand (Soeter & Kindt, 2010, 2011a, 2011b). Recently, we even observed a double dissociation between the emotional (startle potentiation) and cognitive expression (US-expectancies, SCR) of fear memory (Sevenster et al., 2012a). Thus, in line with our previous studies, we found a dissociation between the startle potentiation and SCR (Sevenster et al., 2012a; Soeter & Kindt, 2010, 2011a, 2011b). SCR again shows a pattern closely associated with contingency learning and seems to reflect current states of arousal. The startle response acts relatively independently from declarative knowledge and seems to reflect a more automatic expression of fear memory. Instead of one underlying learning mechanism, affective and cognitive learning seem to originate from distinct learning systems. Higher level cognitive processes serve the detection of stimulus events in the environment and the relationship between these events, whereas automatic lower level processes regulate the transfer of affective valence from the US to the CS. This observation supports a dual process account of fear learning.

The dual process account of fear learning is in line with the idea that separate response systems rely on different neural circuits. Both animal research and patient studies suggest that different brain areas are involved in the expression of emotional memory (expressed through performance) and declarative memory (conscious recollection of events) (Squire, 1992, 2004). During affective learning a previously innocuous stimulus becomes capable of activating the brain’s subcortical defense system. The amygdala mediates learning via direct projections to the brainstem (LeDoux, 2000). Potentiation of the startle response, an amygdala-initiated response (Davis, 2000; Walker & Davis, 2002), is considered the most reliable index of the defense system. In contrast, memory for the consciously acquired association between the conditioned stimulus and its aversive consequence requires the hippocampus (Hunsaker & Kesner, 2013). The involvement of the hippocampus in processing the contingency at a cognitive level is supported by studies that distinguished between trace and delay conditioning: both patients with damage to the hippocampal formation (Clark & Squire, 1998) and healthy unaware subjects (Weike et al., 2007) failed to acquire eye-blink and startle trace conditioning, respectively. Trace conditioning is considered to be hippocampal dependent as it requires conscious knowledge of
the CS-US relationship. The amnesic patients and unaware subjects did, however, acquire delay conditioning at a normal rate. Delay conditioning is supposed to take place automatically and does not require the ability of verbalizing the contingencies. Crucially, during both forms of conditioning the patients and subjects were unaware of the CS-US contingencies, suggesting that the hippocampus subserves the conscious recollection of associations. Also brain imaging studies demonstrate the involvement of the hippocampus in trace conditioning, lending further support for the hypothesis that declarative knowledge depends on the hippocampal complex (Büchel, Morris, Dolan, & Friston, 1998; Knight et al., 2004). Even though the neural underpinnings of the emotional expression (startle response) and cognitive expression (declarative knowledge) of associative fear memory have been extensively investigated, the neural substrates for SCR remain relatively unknown (Critchley, Elliott, Mathias, & Dolan, 2000). Preliminary animal studies show that sympathetic arousal has its origins in the hypothalamus. Via subcortical projections, the hypothalamus innervates the sweat glands (Boucsein, 1992; Wang, 1964). Furthermore, imaging studies associate differential SCR during fear acquisition with cortical and subcortical regions, such as increased activity in the anterior cingulate cortex (ACC) (Büchel et al., 1998) and de-activation in the ventromedial prefrontal cortex (vmPFC) (Milad et al., 2007). Furthermore, the acquisition of differential SCR has been shown to correlate with amygdala activity (Carter, O’Doherty, Seymour, Koch, & Dolan, 2006), suggesting a modulatory role of the amygdala in generation of SCR. Involvement of the amygdala, which is typically considered to be the fear center of the brain, is remarkable since SCR does not seem a specific index for fear. Recently, it has been proposed that the role of the amygdala mediates a broader spectrum of functions. Suggested examples are the involvement in updating value representations such as stimulus-reward learning (Baxter & Murray, 2002; Morrison & Salzman, 2010) and detecting stimulus relevance (Sander, Grafman, & Zalla, 2003). Consequently, not only the startle response but also SCR might be associated with amygdala activity. Yet, in light of the structural and functional complexity of the amygdala, different neural networks that both do involve the amygdala might underlie the two physiological measures. Startle potentiation, an automatic defensive reflex, is an amygdala-initiated response and relies on amygdalar-subcortical interactions (Davis, 2006), whereas electrodermal conditioning may be modulated by but does not originate from the amygdala.
Even though instructed extinction did not immediately affect the startle response, it did enhance the speed of extinction learning. This suggests that lower level automatic and higher cognitive processes are not entirely separate. Once a fear memory is established, the affective component of fear learning guarantees that an individual automatically prepares for action in the face of danger. It is therefore adaptive that affective learning is easily acquired, whereas its subsequent extinction is harder to establish. However, as we observe in the present study, learning at a cognitive level (i.e., the instructions on CS-US contingencies) does influence affective learning eventually (i.e., increase in speed of extinction of the startle response). Although it has been proposed that bidirectional connections between the amygdala, hippocampus and prefrontal cortex (PFC) could lie at the basis of the interaction between emotion and cognition (Salzman & Fusi, 2010), mechanisms of this type of emotion regulation remain poorly understood.

In sum, the results of the present study support past findings of a dissociation between the affective and cognitive expression of fear memory. Both expectancy ratings and SCR closely followed the experimental contingencies. The verbal instructions did not have an immediate reducing effect on the startle potentiation, nor did they prevent the return of fear after reinstatement shocks. Hence, affective learning, unlike expectancy learning, is not directly susceptible to extinction instructions, although they enhance speed of extinction. This suggests that, while not directly affected by higher cognitive processes, the emotional component of fear memory is eventually influenced by cognition. A challenge for future research is to unravel the multiple processes of fear learning and memory and their complex interplay.