Pupil dilation and skin conductance as measures of prediction error in aversive learning

Stemerding, L.E.; van Ast, V.A.; Gerlicher, A.M.V.; Kindt, M.

Citation for published version (APA):

Pupil dilation and skin conductance as measures of prediction error in aversive learning

Lotte E. Stemerding [*], Vanessa A. van Ast, Anna M.V. Gerlicher, Merel Kindt [**]

Department of Clinical Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT, Amsterdam, the Netherlands

A B S T R A C T

Notwithstanding the success of CBT, it is relatively unknown how individuals can better profit from corrective learning experiences. Various theories postulate that prediction errors – the difference between what is occurring and what is expected – are the driving force of associative (re)learning. While prediction errors are typically operationalized as violations of cognitive outcome expectancies, direct physiological indices of prediction errors could capture potentially more essential automatic and emotional processes in associative learning. Although physiological responses have previously been suggested to reflect prediction errors, it remains elusive if these measures actually predict changes in subsequent conditioned responding. In three fear-conditioning experiments, we compared pupil dilation and skin conductance responses to unexpected outcomes – unconditioned stimulus (US) presentations or omissions – with expected outcomes, and tested whether outcome responses predicted actual changes in subsequent conditioned responding. We found evidence for increased physiological responses to unexpected outcomes, but the results were inconsistent across experiments. Furthermore, only pupil responses to US presentations consistently predicted an increase in conditioned responding, making it difficult to reconcile our findings with associative learning models. Both pupil dilation and skin conductance can thus index unexpected outcomes, but the relationship of these responses to future learning is not evident and requires further investigation.

1. Introduction

Decades of Pavlovian fear-conditioning research epitomizes how basic science has significantly advanced our understanding and treatment of fear and anxiety disorders (Craske et al., 2018; Kindt, 2014). In particular, fear extinction research in animals and humans has resulted in the development of exposure treatment, which currently is one of the most effective intervention techniques to put irrational fears at bay. Exposure therapy typically involves repeated confrontation with fear-associated stimuli (i.e., objects, situations, interoceptive cues, or memories) in the absence of the anticipated outcome. Notwithstanding the progress that has been made in the effectiveness of exposure therapy, it is relatively unknown how individuals could better profit from these corrective learning experiences. Various theories postulate that prediction errors – i.e., the difference between what is occurring and what is expected – are the driving force of associative (re)learning (Pearce & Hall, 1980; Rescorla & Wagner, 1972; Sutton & Barto, 1987). While the notion of prediction error has been recognized for decades, its

contribution to treatment success has only more recently been subjected to empirical scrutiny in the context of mitigating the expression of fear memories through extinction learning (Craske et al., 2014) or disrupting fear memory reconsolidation (Sevenster et al., 2012, 2013). Further investigation into the critical role of prediction errors in changing fear memory expression is complicated by the lack of reliable indices of prediction errors at the moment of occurrence. In the current study we not only aim to validate previously proposed measures of prediction errors in a Pavlovian fear-conditioning paradigm, but more importantly, we investigate whether the degree of observed prediction error can predict subsequent associative learning.

One potential index of prediction errors in Pavlovian conditioning paradigms are changes in the extent to which an outcome is expected. If participants predict the occurrence of an outcome during the presentation of a conditioned stimulus (CS), then prediction errors can be inferred from changes in outcome expectation from one trial to the next (Sevenster et al., 2013). However, in most fear-conditioning paradigms, outcome expectancies exclusively capture explicit contingency learning

(related to the occurrence of the US), and are not sensitive to other factors such as the timing, valence or motivational significance of the unconditioned stimulus, which may be equally critical to the magnitude of the prediction error (Laurent et al., 2018; Sutton & Barto, 1987). Moreover, changes in expectancy ratings allow only for a retrospective evaluation of prediction error occurrence. These measures are thus indirect and do not capture prediction errors at the very moment they occur, which is when the outcome (US) is either presented or omitted (Rescorla & Wagner, 1972). It has previously been proposed that prediction errors can alternatively be indexed using physiological measures during outcome responding (responding to the US or the omission thereof; e.g., Willems & Vervliet, 2021), and here we aim to corroborate and further explore these findings. Crucially, if physiological outcome responses reflect prediction errors in associative learning, they should predict future conditioned responding. In the current study we will – in addition to the traditional conditioned response on CSs – measure outcome responses in a Pavlovian conditioning task, which enables us to investigate the relationship between observed prediction errors and changes in future conditioned responding.

An extensive body of literature has identified potential indices of prediction error in reinforcement learning, predominantly using pupil dilation measurements. Changes in pupil dilation are associated with uncertainty in decision making tasks, although pupil dilation has both been reported to decrease (Lavín et al., 2014; Preuschoff et al., 2011) and increase (Satterthwaite et al., 2007) when the outcome of the decision is uncertain compared to when it is certain. Furthermore, pupil dilation was found to increase during the observation of a surprising or unexpected outcome in associative learning (Kloosterman et al., 2015; Lavín et al., 2014; Nassar et al., 2012; O'Reilly et al., 2013; Preuschoff et al., 2011). Error monitoring, arguably closely related to surprise, has likewise been related to an increase in pupil dilation in decision-making studies, with unexpected errors eliciting more pupil dilation than expected errors (Braem et al., 2015; Colizoli et al., 2018; Urai et al., 2017). Together, these studies show that pupil dilation can index surprise and error-detection in reward learning. Pupil dilation measurements of outcome responding in aversive learning are scarce, although Browning et al. (2015) found that pupil dilation can track surprising outcomes in an aversive decision making task. Further, changes in pupil dilation under constant lumination have been linked to noradrenaline (NA) activity in the locus coeruleus (Joshi et al., 2016), which has also been implicated in the signaling of negative prediction errors during extinction (Iordanova et al., 2021). We therefore aim to evaluate whether pupil dilation responses to unexpected outcomes may reflect prediction errors in an aversive learning task.

The second candidate physiological readout of prediction errors in aversive learning is skin conductance responses (SCRs). Studies that investigated SCRs to expected versus unexpected US presentations show that SCRs at outcome are significantly reduced when the US is fully predicted compared to when it is only 50% predicted (Dunsmoor et al., 2008) or completely unpredicted (Knight et al., 2010, 2011). In one study, SCRs to the US were inversely related to US expectancy ratings, indicating that the more the US is expected, the lower the SCR to its occurrence, which is in line with what would be expected from a prediction error signal (Knight et al., 2010). Thus, even though the experience of a US itself triggers a physiological response, this response appears to be moderated by the predictability of the US, an effect known as unconditioned response diminution (Goodman et al., 2018). Moreover, studies investigating SCRs to US omissions show that unexpected US omissions (the non-reinforced trials of a 50% reinforced CS+) generate larger SCRs than expected US omissions (Spoormaker et al., 2011, 2012; Willems & Vervliet, 2021). Interestingly, Willems and Vervliet (2021) found that this effect was modulated by the US intensity the participants expected: Omissions of USs with a higher expected intensity gave rise to larger SCRs than omissions of USs with lower expected intensities. These studies show that SCRs may be used to measure unexpected US omission responses, although others have contested this

(Bach & Friston, 2012). Here we aim to test whether SCRs can reliably index prediction errors in an aversive learning task, and to further investigate whether SCRs at outcome relate to future learning.

In sum, both skin conductance and pupil responses have the ability to index unexpected or surprising outcomes. While increased physiological responding to unexpected outcomes would certainly be expected from a prediction error-like signal, a critical test would be to investigate whether the measured outcome responses relate to future learning. If physiological outcome responses indeed reflect prediction errors, then their magnitude should relate to a change in conditioned responding on a next trial (Pearce & Hall, 1980; Rescorla & Wagner, 1972). Specifically, one would expect that conditioned responding increases after unexpected US presentations and decreases after unexpected US omissions. In contrast with previous studies (Willems & Vervliet, 2021), measuring responses to expected versus unexpected outcomes in a conditioning paradigm enables us to test the relationship between outcome responses and changes in conditioned responding.

In three fear-conditioning experiments we evaluated whether pupil dilation and SCRs can reflect prediction errors in aversive learning. We designed three experiments that ensured high variance in prediction error occurrence based on model simulations (see Supplementary Fig. 1 for simulated data), and we predicted that pupil and SCRs would be larger for unexpected US presentations and omissions than expected ones. Further, to establish whether SCR and pupil responses in these experiments can predict learning, we investigated the relationship between outcome responses and changes in CS responding on a next trial of the same condition. The designs of the three experiments differed slightly: In Experiment 1 we manipulated the reinforcement ratio of three CSs and compared responses to CS outcomes that were not fully predictable (50% reinforcement) with responses to outcomes that were fully predictable (0% and 100% reinforcement). Due to a lack of expectancy ratings, we could only infer CS-US contingency knowledge from a post-acquisition questionnaire, and some participants may not have been certain about the contingencies throughout the experiment which would undermine our manipulation (i.e., participants may still perceive the US following a fully reinforced CS as unexpected). Furthermore, it is unclear what type of predictions are made when a subject is presented with a 50% reinforced CS. Instead of making specific predictions regarding the occurrence of the US, the uncertainty itself may become predictable and the subject may simultaneously predict the occurrence and the omission of the US (for a more extensive discussion of this topic see Tronson, 2020). In Experiment 2 we addressed this issue and changed the contingencies only a few times across trials (e.g., after 6 trials), rather than every other trial. In Experiment 3 we aimed to verify the findings of Experiment 1, while including trial-by-trial expectancy ratings to obtain better insight into the participants' explicit contingency knowledge and how expectancy ratings relate to our proposed physiological measures of prediction error.

2. General methods

2.1. Preregistrations

Experiment 1 was exploratory and was not preregistered. Experiment 2 and 3 were preregistered at the OSF (https://osf.io/y9qmb and https://osf.io/6zjvp).

2.2. Participants

Participants were healthy volunteers with normal or corrected to normal vision (see further demographics for each experiment in the respective methods section). The three experiments were approved by the ethics board of the University of Amsterdam. All participants signed informed consent after being informed about the procedure and were reimbursed with course-credit or 10 Euros/hour for their participation.

2.3. Materials and measures

2.3.1. Conditioned stimuli

Conditioned stimuli (CSs) were three different geometrical shapes (square, triangle, hexagon) with different colors (blue, green, purple), that were presented on a grey background in the upper-middle part of the screen (Fig. 1). The mean luminosity for each stimulus and the background was the same. We manipulated the US probability for each CS. One CS was followed by the US 0% (CS_0), one 50% (CS_{50}) and one 100% (CS_{100}) of the trials. The assignment of each geometrical shape to each US probability was randomized and counterbalanced across participants. Each CS was shown on the screen for 6.5s, and the US, if presented, occurred at offset. Inter trial intervals ranged from 8.5 to 13.5s (mean 11s), during which a fixation cross was presented. Stimulus presentation was semi-randomized, such that no CS was presented more than two times in a row.

2.3.2. Unconditioned stimulus

The unconditioned stimulus (US) consisted of a brief electrical stimulus applied to the top of the left wrist. The stimulus was delivered by a Digitimer DS71 (Welwyn Garden City, UK) through two Ag/AgCl electrodes of 20 by 25 mm with a fixed inter-electrode distance of 45 mm, and for a duration of 2 ms. The intensity of the stimulus was individually determined to be clearly uncomfortable but not painful. Participants first received the lowest stimulus (1 mA), after which the intensity was increased step-by-step with 2–4 mA at the time. Participants were instructed to say "stop" when they felt the stimulus was truly uncomfortable, after which they rated the perceived intensity of the stimulus on a scale ranging from 0 ("I barely felt anything") to 10 ("This is the most uncomfortable stimulus I can imagine to be applied through this electrode"). If participants reached the maximum intensity of 70 mA and they did not rate this as at least 7, they were excluded from participating in the experiment.

2.3.3. Pupillometry

Pupil dilation was recorded with a Tobii Nano Pro eye tracker (Tobii Pro AB, Stockholm, Sweden) using a sampling rate of 60 Hz. To minimize movement, participants kept their head in a chin and forehead rest. Raw pupil data was preprocessed using MATLAB version 2018b (The MathWorks Inc, 2018). Missing samples were identified and samples 100 ms surrounding missing samples were removed. These data points were replaced by linear interpolation. Trials with more than 50% missing data during either baseline or the trial epoch were excluded from analyses. Missing trials were replaced by linear interpolation within each US probability. Participants with more than 33% missing trials in one or more of the three CS conditions were excluded from the entire analysis (Visser et al., 2013). Data were then filtered using a 3rd order Butterworth filter with a cutoff frequency of 6.5 Hz. Epochs of 0–6500 ms after CS onset (CS responding) and 0–3000 ms after US onset (outcome responding) were taken from the continuous data. Epochs for outcome responding were restricted to 3000 ms as timeline plots showed shorter response latencies for outcome responses and effects of surprise during outcome responding have been found within this timeframe (Browning et al., 2015). All epochs were baseline corrected by pupil dilation averaged over 500 ms prior to CS or US onset. By design, the baseline for US epochs overlaps with the last 500 ms of CS presentation. This may be problematic due to ceiling effects for the CS_{100} outcome responses, but we checked how many responses were close to (>95%) the participant's maximum response and found no evidence of ceiling effects. The peak value of each baseline corrected epoch was taken as index of pupil dilation for that trial.

2.3.4. Skin conductance measurements

Skin conductance was measured through two 16×20 mm Ag/AgCl electrodes attached to the medial phalanx surfaces of the index and middle finger of the left hand. Skin conductance was recorded using the software program VSRRP98 and sampled at 1000 Hz. The raw skin conductance signal was preprocessed using MATLAB version 2018b (The MathWorks Inc, 2018). The signal was digitized and filtered using a 1st order Butterworth filter with a cutoff frequency of 1 Hz (Boucsein et al., 2012). In contrast with the preregistrations, we have applied a through-to-peak hand scoring approach for the skin conductance data as this minimizes potential effects of CS responding on outcome responses. Using a custom made analysis script, we identified the first SCR onset (local minimum) in a 900–4000 ms window post stimulus onset for both CS and outcome responses (Sjouwerman & Lonsdorf, 2019). SCRs were then calculated as the difference between the first local minimum and the first subsequent peak (for examples see Supplementary Fig. 3). Data were scored blind to both condition and response type (outcome or CS). For CS responses, all SCRs smaller than 0.02 μS were scored as zero (2–4% of the data). For outcomes we included all responses regardless of their size as it is unknown if this cutoff is appropriate for outcome responses. The results did not change in either of the experiments if we did score outcome responses smaller than 0.02 μS as zero. Across all experiments, six participants were excluded due to recording errors (3 in Exp1, 1 in Exp2, and 2 in Exp3). No further participants were excluded based on SCR criteria.[1]

2.4. General procedure

All experiments consisted of a single session lasting 30 minutes to 1 hour. Participants filled out trait anxiety (STAI-T; Spielberger, 1983) and anxiety sensitivity (ASI; Peterson & Reiss, 1992) questionnaires before the start of the experiment. Trait anxiety may affect fear conditioning, and especially safety learning (Browning et al., 2015; Gazendam et al., 2013) and anxiety sensitivity may affect the response to the US. Both ASI and STAI-T scores did not differ between experiments (see Supplementary Results). Skin conductance and electrical stimulus electrodes were attached to the left hand and arm respectively, and the electrical stimulus was calibrated (see *Unconditioned Stimulus* section). The eye tracker was calibrated using a 6-point calibration. For exact procedural descriptions of the experimental tasks, see methods per experiment.

2.5. Statistical analyses

All data were statistically analyzed in RStudio (version 1.3.1093). For each experiment we first analyzed responses to the CS using a repeated measures ANOVA with the three US Probabilities (CS_0, CS_{50}, CS_{100}) as within-subjects factors (package: *rstatix*, function: *anova_test*, sum of squares type II). Depending on the number of trials in the experiment, trials were averaged across either the entire experiment (Exp1), or across one or two phases consisting of 12 trials each (Exp2 and Exp3). The factor Phase was added to the rmANOVA. Outcome responding was assessed for US presentations and omissions separately because US presentations trigger physiological responding. Only expected outcomes (CS_{100}/CS_0) that matched in trial number with unexpected outcomes (CS_{50}) were included in the analyses (i.e., if CS_{50} trials 2, 3, 5, 7, 8, 10 were reinforced, only CS_{100} trials 2, 3, 5, 7, 8, 10 were included). To test for differences between expected and unexpected outcomes, we performed a repeated measures ANOVA with US Probability (CS_0, CS_{50}, CS_{100}) and Trial (1–6/1-3/1–12, depending on experiment) as within-subject factors. Here, we deviated from the original preregistration. We planned to use a *t*-test on the average of all trials but decided to instead perform an ANOVA including individual trial data (see methods per experiment for the exact trials included), as this allowed us to include more data points in the analyses, as well as to

---

[1] Excluding pupil exclusions from SCR data changed the results for outcome responses in Exp1 and Exp2. See Supplementary Materials for SCR analyses when excluding the same participants that were excluded for pupil data.
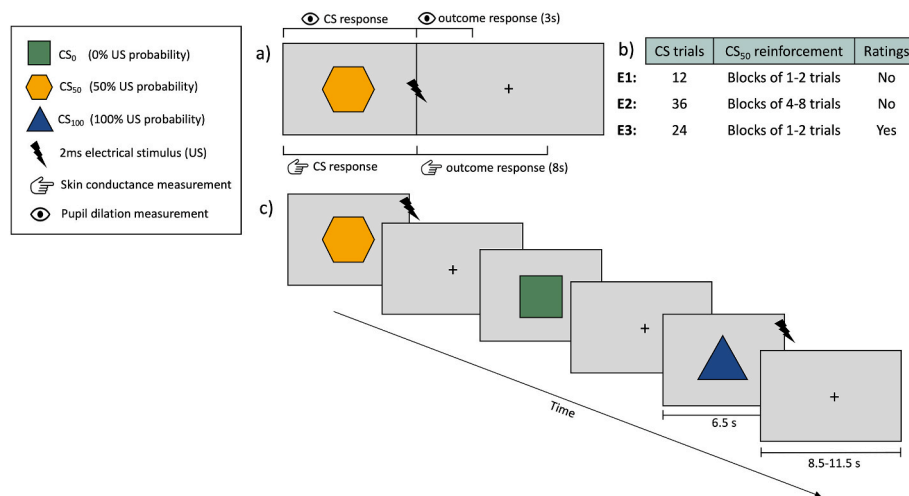
Fig. 1. Experimental design. a) Example of one trial with both pupil and SCR epochs indicated. b) Schematic overview of the differences between the three experiments in terms of number of presentations of each US probability stimulus, the reinforcement of the CS_{50}, and whether participants provided US expectancy ratings or not. c) Three example trials of different US probabilities across time (CS_0, CS_{50}, CS_{100}).

investigate potential effects of time.

Lastly, we investigated whether the magnitude of outcome responding could predict a change in conditioned responding (CR) to the CS on the next trial of the same US Probability. Because the trial-level data are nested within participants, we performed a multilevel model for each experiment using the R-package lme4 (Bates et al., 2015) and the *lmer* function, regressing changes in conditioned responding on outcome responses. The *p*-values for each regressor or interaction were provided by the *summary* function from the R-package lmerTest (Kuznetsova et al., 2017). Only CS_{50} and CS_{100} trials were included in the models because the CS_0 served mainly as expected omission control stimulus and required little updating. As illustrated in Fig. 3, the CR change score was calculated as $CR_{CSt+1} - CR_{CSt}$ within each US probability type and the outcome response at trial t was the main predictor variable. Importantly, responses to unexpected US omissions should reduce responding on a next trial, whereas responses to unexpected US presentations should increase responding. To properly test for this interaction, we pooled all outcome responses together and included an Outcome Response × Reinforcement (US presented, US omitted) interaction in the models. However, the habituation slopes of the outcome responses differed between reinforced and unreinforced trials for pupil data in Experiment 1 and Experiment 2. As the differential effects of habituation on these responses would make the effects unfathomable, we have separated the analyses for the pupil data in these two experiments. Further, due to the randomized CS presentations, some CSs of the same US Probability type would follow each other directly, whereas between others, CSs of a different type were presented. To exclude the possibility that the relation between outcome responding and CS responding is driven by the fact that two CSs occur close to each other in time, we included a factor Distance in the equation. This factor indicated whether the CSs of the same type were presented directly after each other (close) or not (distant). Finding a relationship between outcome responding and a change in conditioned response only for close trials would indicate that the observed effects are not related to learning. Lastly, to control for possible effects of habituation, we also included trial number (the absolute number in the order of presentation) as a predictor. US omission trials were set as the reference category for Reinforcement, thus all beta values for Reinforcement or interactions with Reinforcement can be interpreted as the predicted difference between a presented US versus and omitted US. Distant trials were set as the reference category for Distance. We included a random intercept per subject, resulting in the following model:

CR change ~ US_t × reinforcement × distance + $trial_t$ + (1 | subject)

Raw (unstandardized) data were used. The US response predictor variable was centered within participant (Enders & Tofighi, 2007). Cook's distance (package = influence.ME, function = influence, output of that function used in function cooks.distance) was used to estimate the influence of individual data points on the model outcome. All data points with a Cook's distance larger than $4/n$ were excluded from the model (Nieuwenhuis et al., 2012), which resulted in one data point for SCRs in Exp1, one data point for pupil responses in Exp2, and 8 data points for SCRs in Exp2. Including these data points did not change the results of the model, except for the SCR data in Exp2, where including the data points gives rise to an interaction between US Response and Reinforcement (see Supplementary Results).

3. Experiment 1

3.1. Methods experiment 1

3.1.1. Participants

Forty healthy volunteers (11 male) between 18 and 57 years old (mean ± SD age: 21.8 ± 7.7) participated in this study. The mean US intensity was 17.6 mA (range: 4–55 mA). For pupil data, 10 participants were excluded due to poor pupil data quality (more than 33% missing trials), leaving 30 participants in the main analysis. For SCR data, three participants were excluded due to a technical recording error, leaving 37 participants for statistical analysis.

3.1.2. Procedure and design

During the experimental task, the three conditioned stimuli were each shown 12 times. To ensure contingency knowledge developed similarly for all participants, the reinforcement schedule for the CS_{50} was fixed and the same for all participants (Fig. 2a). Participants were not given any explicit instructions about the conditioned stimuli. After the end of the experiment, participants were asked to rate the believed outcome probability of each stimulus using a slider from 0 to 100.

3.1.3. US probability awareness

For the experimental manipulation to be effective, participants should be aware that the CS_0 and CS_{100} were respectively never and always followed by the US. Therefore, we separately analyzed only participants who learned the outcome probabilities correctly (see Supplementary Results). Probability ratings were classified as correct when participants indicated that the CS_0 was on 0% of the trials followed by the US, the CS_{100} on 100% of the trials, and the CS_{50} between 30% and

Fig. 2. Reinforcement schemes of a) Experiment 1, where the reinforcement of the CS50 was fixed across participants. b) Experiment 2, where only the marked trials that occurred right after a contingency switch were included. Half of the participants started with unreinforced trials (as portrayed here) and half of the participants started with reinforced trials. c) Experiment 3, where the reinforcement of the CS50 was semi-random with the requirement that no more than two trials in a row were reinforced or unreinforced.
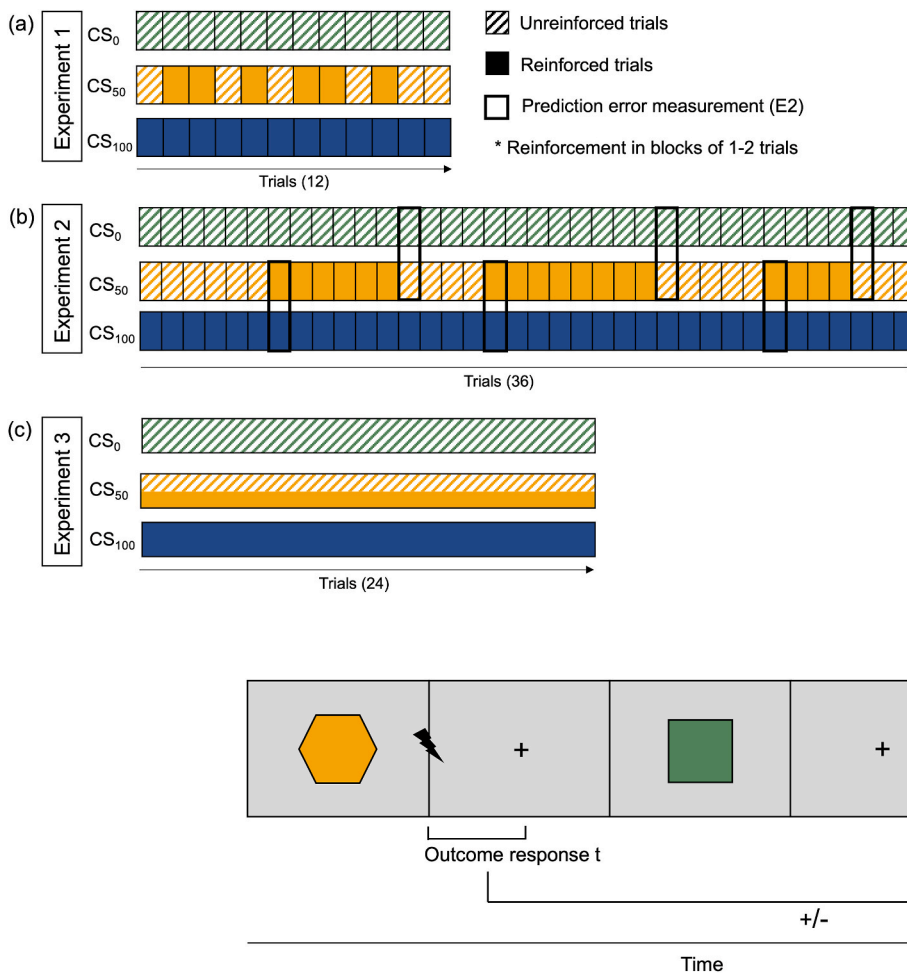

Fig. 3. Schematic overview of the multilevel approach to test the relationship between outcome responses and subsequent changes in CS responding. In this example, the distance factor would be "distant" because there is one different US Probability type presented in between.

70%. If the results from this selected sample differed from the results based on the entire sample this has been indicated in the main text.

3.2. Results experiment 1

3.2.1. Responding to the conditioned stimuli

Pupil responses (Fig. 4a). A repeated measures ANOVA with US Probability (CS_0, CS_{50}, CS_{100}; average of the 12 trials) as within-subject factor indicated a significant main effect of US Probability ($F_{(2,58)}$ = 8.17, p < .001, η_p^2 = 0.22). Planned comparisons showed that pupil responding to the CS_{100} was higher than the CS_0 ($t_{(29)}$ = 3.64, p = .001). Further, pupil responses to the CS_{50} were significantly larger than to the CS_0 ($t_{(29)}$ = 3.04, p = .005). Responding to the CS_{50} did not significantly differ from the CS_{100} ($t_{(29)}$ = 0.77, p = .449). Thus, across trials, participants showed stronger pupil responses to both the CS_{50} and the CS_{100} compared to the CS_0, which is indicative of fear learning.

Skin conductance responses (Fig. 4d). We found a significant effect of US Probability ($F_{(1.4,49.2)}$ = 10.41, p < .001, η_p^2 = 0.22), and planned comparisons showed that SCRs to the CS_{100} were on average higher than to both the CS_{50} ($t_{(36)}$ = 2.88, p = .007) and the CS_0 ($t_{(36)}$ = 3.58, p = .001). Further, SCRs to the CS_{50} were larger than to the CS_0 ($t_{(36)}$ = 2.53, p = .016), indicating that anticipatory SCRs developed for both the CS_{50} and CS_{100}, with stronger responding to a consistently reinforced CS.

3.2.2. Responding to the outcomes

Pupil responses (Fig. 5a). To test whether responding to

unexpected outcomes (following the CS_{50}) would be larger than to expected outcomes (following CS_0 and CS_{100}), we performed two separate US Probability (CS_{50}, CS_0/CS_{50}, CS_{100}) × Trial (1–6) repeated measures ANOVAs on the responses to US omissions and presentations respectively. In contrast with our hypothesis, we found no significant main effect of US Probability on responding to US presentations ($F_{(1,29)}$ = 0.05, p = .822, η_p^2 < 0.01). In the sample including only participants who learned the US probabilities we found a significant US Probability × Trial interaction, and follow-up analyses demonstrated that US responding was significantly larger to unexpected US presentations only on the 5th experimental trial (see Supplementary Results). For US omission responses we did not find a main effect of US Probability either ($F_{(1,29)}$ = 3.39, p = .076, η_p^2 = 0.11). These results suggest that the unexpectedness of the outcome does not significantly affect the outcome response across the entire experiment.

Skin conductance responses (Fig. 5d). In line with our predictions, we found a significant main effect of US Probability for responses to US presentations ($F_{(1,36)}$ = 4.72, p = .036, η_p^2 = 0.12). For US omissions responses we did not observe the expected main effect of US Probability ($F_{(1,36)}$ = 2.21, p = .146, η_p^2 = 0.06). These results suggest that the unexpectedness of the outcome significantly increases SCRs to US presentations, which may be driven by prediction error occurrence. However, we did not find this effect for US omissions.

3.2.3. The relationship between CS and outcome responding

If outcome responses indeed reflect prediction errors, then the magnitude of these responses should be predictive of a change in
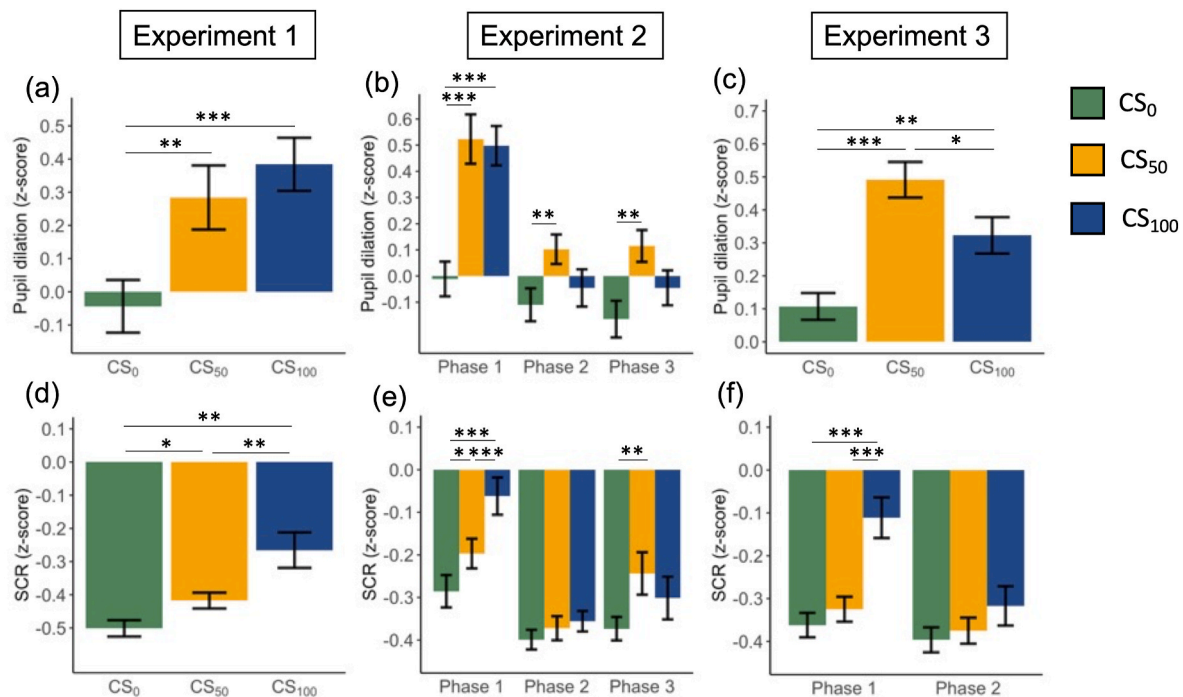
Fig. 4. Standardized pupil dilation (a–c) and SCR (d–f) responses to CS presentations. (a) Pupil responses during CS_0, CS_{50} and CS_{100} presentation in Experiment 1, averaged across all 12 trials. (b) Pupil responses during CS_0, CS_{50} and CS_{100} presentation in Experiment 2, visualized per phase of 12 trials due to a significant Phase × US probability interaction. (c) Pupil responses during CS_0, CS_{50} and CS_{100} presentation in Experiment 3, averaged across all 24 trials. (d) Skin conductance responses to CS_0, CS_{50} and CS_{100} presentation in Experiment 1, averaged across all 12 trials. (e) Skin conductance responses to CS_0, CS_{50} and CS_{100} presentation in Experiment 2, visualized per phase of 12 trials due to a significant Phase × US probability interaction. (f) Skin conductance responses to CS_0, CS_{50} and CS_{100} presentation in Experiment 3, visualized per phase of 12 trials due to a significant Phase × US probability interaction. Error bars depict standard error of the mean. ***p < .001, **p < .01, *p < .05.

physiological responding on a next trial of the same US probability. Specifically, CS responding should increase after unexpected US presentations, and decrease after unexpected US omissions. We tested this hypothesis using a multilevel model and included all experimental trials.

Pupil responses. Parameter estimates from the multilevel model for US presentations showed a significant effect of Outcome Response (β = .31, t(415) = 3.83, p < .001), indicating that a stronger response was associated with a larger subsequent increase in conditioned responding. For US omission responses, there were no significant predictors in the model (Outcome Response: β = 0.17, t(539) = 0.81, p = .422), showing that there is no relationship between responses to US omissions and a change in conditioned responding. Notably, the beta estimate for this factor is positive which is in contrast with our hypothesis that responses to US omissions would predict a decrease in conditioned responding.

Skin conductance responses. Parameter estimates for the model on SCR data showed that the expected Outcome Response × Reinforcement interaction was not significant (β = .24, t(805) = 1.84, p = .066). Rerunning the model without the Outcome Response × Reinforcement interaction showed no further significant effects. These results do not confirm our hypothesis, as we found no evidence for an effect of outcome responses on conditioned responding on a subsequent trial of the same US Reinforcement.

4. Experiment 2

In Experiment 1 we observed that unexpected outcomes were indeed associated with larger responses, but not all effects were statistically significant. We assumed that the 50% occurrence of the US may have led to a form of expected uncertainty (Soltani & Izquierdo, 2019), where participants generated predictions such as "I may or may not receive an electrical stimulus", causing either outcome to be expected. To strengthen prediction errors, we designed an experiment similar to

Experiment 1, but where the CS_{50} would be reinforced for a series of 4–8 trials, rather than 1–2 trials, and then unreinforced for a series of 4–8 trials (Fig. 2b). Simulated data based on a Rescorla Wagner learning rule shows that prediction errors are larger when more sustained predictions can be made and reach a maximum directly after a switch in reinforcement (Supplementary Fig. 2). We therefore only selected switch trials (trials that followed a reinforcement switch) for analyses (Fig. 2b). The CS_0 and CS_{100} again served as control stimuli.

4.1. Methods experiment 2

4.1.1. Participants

Forty healthy volunteers (11 male) between 18 and 24 years old (mean ± SD age: 19.8 ± 1.6) participated in this study. The mean US intensity was 17.4 mA (range: 6–48 mA). For pupil analyses, 8 people were excluded due to poor pupil data quality (more than 33% of trials missing for any US Probability) leaving a total sample of 32 participants. For SCR analyses, one participant was excluded due to a technical recording error, leaving a total sample of 39 participants.

4.1.2. Procedure and design

To promote certainty about the CS_0 and CS_{100} stimuli, participants were told that these stimuli would respectively never and always be followed by the US. For the CS_{50} they were instructed to learn to predict when the US would occur. During the experiment, each stimulus was presented 36 times. See Fig. 2b for the exact pattern of reinforcement of the CS_{50}. This pattern was the same for all participants, but half of the participants started with reinforced trials, and the other half with unreinforced trials. After the experiment, participants indicated for each CS the percentage of total trials they believed were followed by an electrical stimulus, as well as how many times they believed the reinforcement of the CS_{50} stimulus switched.
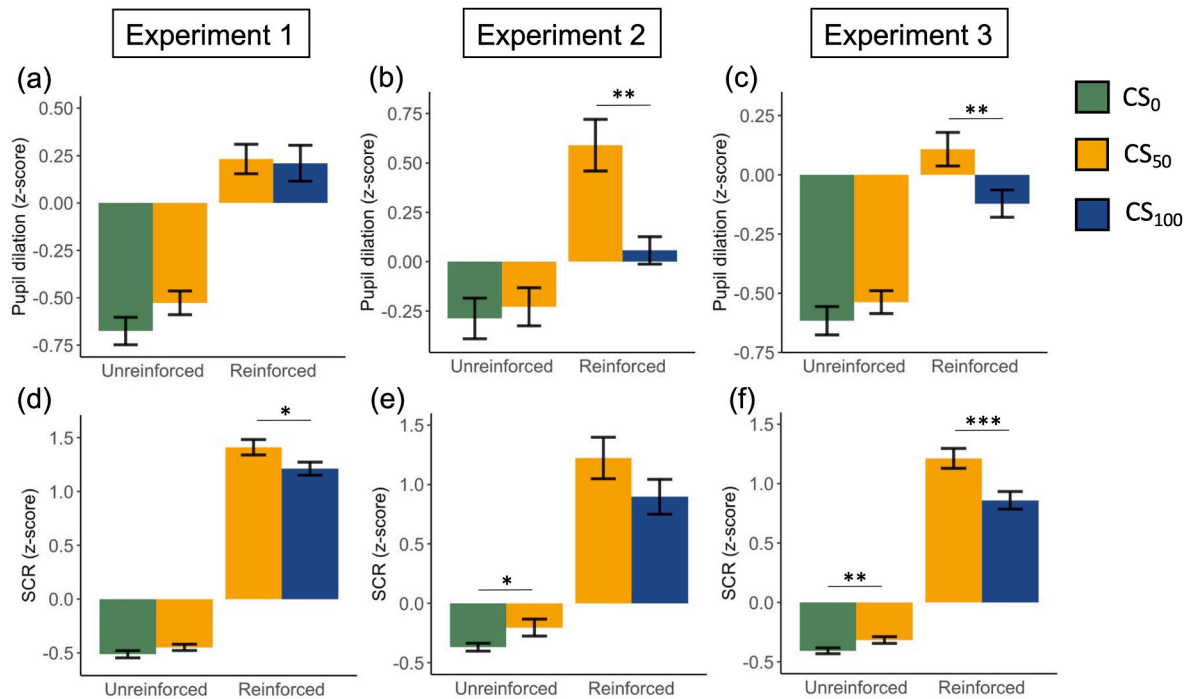
Fig. 5. Standardized pupil dilation (a–c) and SCR (d–f) responses to outcomes. (a) Pupil responses to expected (CS_0, CS_{100}) versus unexpected US (CS_{50}) presentations and omissions in Experiment 1, averaged across 6 trials. (b) Pupil responses to expected versus unexpected US presentations and omissions in Experiment 2, averaged across 3 trials. (c) Pupil responses to expected versus unexpected US presentations and omissions in Experiment 3, averaged across 12 trials. (d) Skin conductance responses to expected (CS_0, CS_{100}) versus unexpected US (CS_{50}) presentations and omissions in Experiment 1, averaged across 6 trials. (e) Skin conductance responses to expected versus unexpected US presentations and omissions in Experiment 2, averaged across 3 trials. (f) Skin conductance responses to expected versus unexpected US presentations and omissions in Experiment 3, averaged across 12 trials. Error bars depict standard error of the mean. ***$p < .001$, **$p < .01$, *$p < .05$.

4.2. Results experiment 2

4.2.1. Responding to the conditioned stimulus

Pupil responses (Fig. 4b). Due to the increased length of the experiment, we analyzed CS responding in three phases, averaging across 12 trials of each US Probability in each phase. The US Probability (CS_0, CS_{50}, CS_{100}) × Phase (first, second, third; average of 12 trials per phase) interaction was significant ($F(4,124) = 4.93$, $p = .003$, $\eta_p^2 = 0.04$). Follow-up analyses showed that the effect of US Probability was significant in the first ($F(2,62) = 20.62$, $p < .001$, $\eta_p^2 = 0.24$), second ($F(2,62) = 3.52$, $p = .036$, $\eta_p^2 = 0.06$), and third phase ($F(2,62) = 4.42$, $p = .016$, $\eta_p^2 = 0.09$). Planned comparisons per phase showed that in the first phase, pupil responses to the CS_{100} ($t(31) = 6.62$, $p < .001$) and the CS_{50} ($t(31) = 4.93$, $p < .001$) were both significantly larger than to the CS_0. In both the second ($t(31) = 2.78$, $p = .009$) and third ($t(31) = 3.31$, $p = .002$) phase only CS_{50} responding was still larger than CS_0 responding, while differences between the CS_{100} and the CS_0 disappeared (both $ps >$ .243). These results demonstrate that pupil responding to a consistently reinforced CS reduces as learning progresses, while responding to a CS with higher uncertainty regarding the outcome remains elevated.

Skin conductance responses (Fig. 4e). A US Probability (CS_0, CS_{50}, CS_{100}) × Phase (first, second, third) rmANOVA showed a significant main effect of US Probability ($F(2,76) = 14.95$, $p < .001$, $\eta_p^2 = 0.04$) and Phase ($F(1.6,62.2) = , p < .001$, $\eta_p^2 = 0.11$), and a US Probability × Phase interaction ($F(3.2,119.7) = 5.27$, $p = .002$, $\eta_p^2 = 0.03$). Follow-up analyses showed that the main effect of US Probability was only significant in Phase 1 ($F(1.7,64.9) = 15.95$, $p < .001$, $\eta_p^2 = 0.13$) and Phase 3 ($F(2,76) = 4.88$, $p = .001$, $\eta_p^2 = 0.04$). Planned comparisons per phase showed significantly larger SCRs to the CS_{100} than to the CS_0 ($t(38) = 5.07$, $p < .001$) and to the CS_{50} ($t(38) = 4.42$, $p < .001$) in Phase 1. Further, in Phase 1 we also found lager SCRs to the CS_{50} compared to the CS_0 ($t(38) = 2.04$, $p = .049$). In Phase 3 we only found larger responses

to the CS_{50} compared to the CS_0 ($t(38) = 3.41$, $p = .002$). Thus, while conditioned responding was reduced in Phase 2, these results are indicative of fear learning at the start of the experiment, with sustained responses to the CS_{50} at the end of the experiment.

4.2.2. Responding to the outcomes

Pupil responses (Fig. 5b). Responses to the three outcomes of trials immediately after a reinforcement change (switch trials, Fig. 2b) were included in the analyses in two US Probability (CS_{100}/CS_{50} or CS_{50}/CS_0) x Trial (1–3) repeated measures ANOVAs. Confirming our predictions, we found a significant main effect of US Probability on responding to US presentations ($F(1,31) = 8.00$, $p = .008$, $\eta_p^2 = 0.21$), showing that across the three trials, responses to unexpected US presentations were larger than to expected presentations. Furthermore, the US Probability × Trial interaction was significant ($F(2,62) = 3.28$, $p = .044$, $\eta_p^2 = 0.10$). Planned comparisons showed that responses to unexpected USs were significantly larger than to expected USs on the first ($t(31) = 2.76$, $p = .010$) and second ($t(31) = 2.09$, $p = .045$) trial, but not on the third ($t(31) = 0.37$, $p = .713$), indicating that the unexpectedness of the outcome did no longer increase pupil responding towards the end of the experiment. Analyses of the US omission responses showed neither a main effect of US Probability ($F(1,31) = 2.20$, $p = .148$, $\eta_p^2 = 0.07$), nor a US Probability × Trial interaction ($F(2,62) = 1.70$, $p = .190$, $\eta_p^2 = 0.05$), indicating that the unexpectedness of the omissions did not result in a larger pupil response.

Skin conductance responses (Fig. 5e). In contrast with our hypothesis and the results of Experiment 1, we found only a trend-wise significant main effect of US Probability ($F(1,38) = 3.53$, $p = .068$, $\eta_p^2 = 0.085$) for US presentation responses. We did find a significant main effect of Trial ($F(1.5,58.3) = 4.13$, $p = .030$, $\eta_p^2 = 0.10$), showing that outcome responding decreased over time. Further, for US omission responses we found a significant main effect of US probability ($F(1,38) =$

4.80, p = .035, $\eta_p^2 = 0.11$), showing that US omission responses were larger for unexpected omissions than for expected omissions. Overall, these results show that US omission responses are larger when the omission is unexpected, and reveal some, yet only trend-wise significant, evidence that SCRs to unexpected US presentations are also larger than to expected ones.

4.2.3. The relationship between CS and outcome responding

Pupil responses. In line with the results of Experiment 1, we found a significant effect of Outcome Response on a change in conditioned responding for US presentations ($\beta = 0.16$, t(1216) = 3.50, p < .001). For US omission responses this effect was merely trend-wise significant ($\beta = 0.21$, t(424) = 1.88, p = .061). These data show that, in contrast with our hypothesis, responses to both US presentations and US omission positively relate to an increase in future conditioned responding, yet this effect is only significant for US presentations.

Skin conductance responses. The initial multilevel analysis including all data showed a significant Outcome Response × Reinforcement interaction, but an influencer analysis (Cook's distance >4/n) showed that this result was driven by eight data points (see Supplementary Results for original analysis). When removing these data points from the analysis, the Outcome Response × Reinforcement interaction was only trend-wise significant ($\beta = 0.10$, t(2674) = 1.82, p = .069). Taking the Reinforcement term out of the model showed no main effects of Outcome Response ($\beta = 0.01$, t(2674) = 1.28, p = .202), indicating that there was no overall relationship between outcome responding and a change in conditioned responding. We thus found no convincing evidence for a predictive relationship between outcome responses and conditioned responding, even though the results appear to go into the right direction.

5. Experiment 3

The third experiment served to replicate the effects of Experiment 1 and optimize the certainty of the CS_0 and CS_{100} through clearer instructions and more learning trials. While the results of Experiment 1 were in the expected direction, they were not statistically significant. Including only participants who rated the contingencies correctly slightly improved our results. However, even these participants may have experienced uncertainty regarding the CS_0 and CS_{100} during the experiment itself. Therefore, in the third experiment the participants were explicitly instructed that one stimulus was never followed by the US and one stimulus always. To ensure that the instructions were understood we included a manipulation check by measuring US expectancy ratings on a trial-by-trial basis. Because expectancy ratings have been used to index prediction errors (Sevenster et al., 2013, 2014), we were also interested to learn whether physiological measures of skin conductance due to unexpected outcomes relate to subjective outcome expectancy. Including expectancy ratings further enabled us to compare unexpected outcomes with expected outcomes based on the violations of expectancy ratings, rather than solely based on the manipulation in the design. We therefore exploratively investigated the relationship between outcome responding and expectancy ratings.

5.1. Methods experiment 3

5.1.1. Participants

Forty healthy volunteers (7 male) between 18 and 34 years old (mean ± SD age: 21.5 ± 3.9) participated in this study. The mean US intensity was 16.0 mA (range: 4–60 mA). For pupil analyses, 7 participants were excluded due to poor pupil data quality (more than 33% trial missing for any US Probability), leaving a total sample of 33 participants. For SCR analyses, two participants were excluded due to a technical recording error, leaving a total sample of 38 participants for statistical analysis.

5.1.2. Procedure

The third experiment consisted of two blocks of twelve trials of each US Probability, separated by a 2-min break. We added the break to avoid that the experiment became too strenuous, and that people lost interest. The reinforcement schedule for the CS_{50} was random, with the restriction that the CS_{50} was reinforced or unreinforced not more than twice in a row (Fig. 2c). Participants were instructed that one stimulus was never followed by the US (CS_0), one stimulus always (CS_{100}), and for the third stimulus they had to learn to predict the US occurrence (CS_{50}). Participants rated their US expectancy within the first 5 s of each stimulus presentation on a scale from 1 to 5 (1 = "will certainly not receive an electrical stimulus", 5 = "will certainly receive an electrical stimulus" with "uncertain" as middle point).

5.2. Results experiment 3

5.2.1. CS responding

Pupil responses (Fig. 4c). We analyzed the data in two phases of twelve stimuli. Comparing average pupil responses for each phase between the three US probabilities in a US Probability (CS_0, CS_{50}, CS_{100}) × Phase (first, second; average of 12 trials per phase) rm ANOVA showed a main effect of US Probability (F(2,64) = 21.33, p < .001, $\eta_p^2 = 0.40$), a main effect of Phase (F(1,32) = 64.80, p < .001, $\eta_p^2 = 0.67$), and no US Probability × Phase interaction (F(2,64) = 0.10, p = .907, $\eta_p^2 = 0.00$). Responding to the CS_{100} was significantly larger than to the CS_0 (t(65) = 4.04, p < .001). In contrast with our observations in Experiment 1, pupil responses to the CS_{50} were larger than to both the CS_{100} (t(65) = 3.23, p = .002) and the CS_0 (t(65) = 7.35, p < .001). These observations corroborate our findings from Experiment 2 and suggest that pupil dilation during CS presentation is predominantly driven by uncertainty about the outcome, rather than by outcome expectation.

Skin conductance responses (Fig. 4f). The US Probability (CS_0, CS_{50}, CS_{100}) × Phase (first, second) rm ANOVA gave a significant main effect of US Probability (F(2,74) = 16.59, p < .001, $\eta_p^2 = 0.31$) and Phase (F(1,37) = 14.16, p < .001, $\eta_p^2 = 0.28$), and a US Probability × Phase interaction (F(2,74) = 5.17, p = .008, $\eta_p^2 = 0.12$). Planned comparisons showed that in the first phase, SCRs were significantly higher to the CS_{100} compared to both the CS_0 (t(37) = 5.21, p < .001) and the CS_{50} (t(37) = 4.09, p < .001). Responding to the CS_{50} did not significantly differ from the CS_0 (t(37) = 1.06, p = .296). In the last phase, none of the conditioned responses significantly differed from each other (all ps > .089). These results align with Experiment 1 and suggest that, in experiments in which the 50% reinforcement is random, SCRs respond more strongly to a stimulus with a higher reinforcement rate.

5.2.2. Responding to the outcomes

Pupil responses (Fig. 5c). The US Probability (CS_{50}, CS_{100}) × Trial (1–12) rm ANOVA performed on pupil responses to US presentations showed a main effect of US Probability (F(1,32) = 9.58, p = .004, $\eta_p^2 = 0.23$) indicating that across trials, responses to unexpected US presentations were larger than to expected US presentations. This is in line with results from Experiment 2. We further observed a main effect of Trial (F(7.51, 240.42) = 2.81, p = .006, $\eta_p^2 = 0.08$), indicating that responses to the US decreased throughout the experiment. Responses to US omissions did not differ between the two US Probabilities (F(1,32) = 2.25, p = .143, $\eta_p^2 = 0.07$). These results confirm the findings from Experiment 2 and suggest that unexpected US presentations elicit greater pupil responses than expected ones, but that expectancy did not affect pupil responses to US omissions.

Skin conductance responses (Fig. 5f). In line with our predictions, the main effect of US Probability is significant for responses to US presentations (F(1,37) = 18.68, p < .001, $\eta_p^2 = 0.34$), showing that SCRs to unexpected US presentations are larger than to expected ones. The main effect of Trial is also significant (F(4.0, 149.4) = 4.71, p = .001, $\eta_p^2 = 0.11$), indicating that responding to US presentations habituates over time. Responses to US omissions are also significantly larger for

unexpected US omissions compared to expected ones, as indicated by the main effect of US Probability (F(1,37) = 10.89, p = .001, η_p^2 = 0.23). These results confirm our hypothesis and suggest that outcome SCRs can index unexpected outcomes.

5.2.3. The relationship between CS responding and outcome responding

Pupil responses. The multilevel model for all outcome responses showed a significant Outcome Response × Reinforcement interaction (β = .38, t(1108) = 2.60, p = .009), indicating that the relationship between outcome responses and changes in CS responding differed significantly between reinforced and unreinforced trials. Analyzing the slopes indicated that in case of reinforced outcomes a greater pupil responses to the outcome predicted a greater increase in responding to the CS on the subsequent trial (β = .36, t(1108) = 5.46, p < .001). This relationship was not significant for unreinforced outcomes (β = −.02, t (1108) = 0.13, p = .898). These results confirm our findings from the first two experiments, except that here we find a negative beta value for US omissions.

Skin conductance responses. The Outcome Response × Reinforcement interaction was not significant (β = −.08, t(1701) = 0.93, p = .352), showing that the relationship between SCRs to the CS outcome and a change in SCRs to the CS presentation did not differ between reinforced and unreinforced outcomes. In line with the previous experiments, removing the Outcome Response × Reinforcement interaction from the model showed that Outcome Response did not significantly predict a change in CS responding (β = .004, t(1704) = 0.22, p = .823), indicating that there is no relationship between outcome responding and a change in CS responding in the data.

5.2.4. The relationship between outcome responding and US expectancy ratings

In the third experiment, participants rated their US expectancy on each trial on a scale from 1 (definitely no US) to 5 (definitely US). While in our main analyses we assumed that the CS_{50} outcomes are unexpected and the CS_0 and CS_{100} outcomes are expected, the use of expectancy data allowed us to compare actually unexpected outcomes to actually expected outcomes. For each trial we calculated an absolute expectancy violation value as [actual outcome – expected outcome], resulting in values ranging from 0 (no violation) to 4 (strong violation). We then categorized outcome responses as either "Expected" (0) or "Unexpected" (1 or higher) based on expectancy violations. Comparing actually expected versus unexpected US omissions and presentations in a paired *t*-test confirmed our findings from the main analyses, showing stronger responses to unexpected outcomes. We further explored whether the size of the outcome response was linearly related to the size of the expectancy violation (i.e., largest for strong violations), and found this effect only for pupil responses to US presentations, indicating that these responses most directly reflect expectancy violations. Lastly, we exploratively tested whether an outcome response on a given trial can predict a change in US expectancy from the present to the next trial of the same US probability. We found no evidence for this relationship in either pupil or SCR data (see Supplementary Results for all exploratory analyses including expectancy data).

6. Discussion

The current study aimed to evaluate whether skin conductance and pupil responses to outcomes in an aversive Pavlovian learning task could serve as a direct measure of outcome-driven prediction errors. In a series of experiments designed to maximize prediction error occurrence, we tested whether physiological responses to unexpected outcomes (US presentations and omissions) would be stronger than to expected outcomes. We found some evidence for stronger pupil responses to unexpected US presentations (Exp 2 and 3), and stronger SCRs to unexpected US presentations (Exp 1 and 3) and unexpected US omissions (Exp 2 and 3). While these findings are in line with previous results (Willems &

Vervliet, 2021) – suggesting that especially SCRs may be used to index unexpected outcomes in an aversive learning task – the results were not entirely consistent across the experiments. Further, we found that only the magnitude of pupil responses to US presentations predicts an increase in conditioned responding on a subsequent trial of the same CS. These results do not entirely confirm our hypothesis based on prediction-error driven updating (i.e., conditioned responses increase after unexpected US presentations and decrease after unexpected US omissions), which complicates aligning our findings with associative learning models.

For pupil dilation, we observed stronger responses to unexpected US presentations in the final two experiments, potentially indicative of positive prediction error signaling. In contrast, we found no evidence for stronger pupil dilation to unexpected US omissions in either of the experiments. This is surprising given the body of literature implicating pupil dilation as a prediction error or surprise signal in reinforcement learning (Kloosterman et al., 2015; Lavín et al., 2014; O'Reilly et al., 2013; Preuschoff et al., 2011; Van Slooten et al., 2018). While the direction of the effect for pupil dilation to unexpected US omissions is in line with these studies, the effects are statistically nonsignificant. Notably, Pavlovian conditioning experiments consist of relatively few trials (12–36 here compared to 200–300 in most decision-making studies), and one reason for our nonsignificant results may be that more trials are necessary to observe a robust effect. In fear conditioning however, the use of hundreds of trials is virtually impossible, as strong habituation to the US would occur. There are two alternative explanations for increased pupil responses to unexpected outcomes. First, pupil responses have also been found to reflect reinforcement changes in a dynamic environment (Nassar et al., 2012). Whereas in Experiment 1 and 3 the CS_{50} stimulus was randomly reinforced, in Experiment 2 the reinforcement switches occurred less frequently. In the latter experiment, an unexpected outcome thus signals a change in future reinforcement (i.e., a change in the environment). The larger pupil responses in Experiment 2 could therefore also reflect the detection of an environmental change rather than larger stochastic prediction errors. It should be noted, however, that in the current design the prediction error size (see Supplementary Fig. 2) and the volatility of the environment were manipulated simultaneously. Hence, we can only speculate on which processes or combination of processes drive pupil responses to outcomes. That said, the larger responses to unexpected US presentation in Experiment 3, which constitutes no change in environment, suggest that also smaller prediction errors are captured by the pupil data. Secondly, the larger pupil responses to unexpected US presentations could be driven by higher levels of US-driven arousal, which increases pupil responding as well (Bradley et al., 2008). This would explain the absence of an effect on US omissions. In sum, pupil responses are larger for unexpected US presentations in the last two experiments, but the lack of a similar response for US omissions makes it difficult to conclusively interpret these findings as reflecting prediction errors.

Increased SCR responding to unexpected compared to expected US presentations has previously been found in studies of unconditioned response reduction, showing that responding to the US decreases as it becomes more expected (Dunsmoor et al., 2008; Knight et al., 2010, 2011). These observations may well be explained by a prediction error hypothesis and are in line with our own findings in the first and third experiment. The absence of a strong effect in the second experiment is slightly puzzling, as the prediction errors are largest here. However, only the three switch trials are included in the analyses of the second experiment, which obviously reduced the statistical power. We further observed larger SCRs to unexpected versus expected US omissions in the second and third experiment. These results add to a growing body of literature showing SCRs to unexpected US omissions (Spoormaker et al., 2011, 2012; Willems & Vervliet, 2021). The fact that the first experiment was entirely uninstructed may explain why we did not find larger SCRs to unexpected US omissions here. Indeed, 18 participants incorrectly reported the CS-US contingencies after the experiment yet

excluding them did not change the results. Still, participants who did correctly indicate that the CS_0 was not followed by the US in hindsight may not have entirely trusted the CS_0 as being certainly safe throughout the experiment, which could have reduced potential differences between the CS_0 and CS_{50} conditions. Our findings thus support the suggestion that SCRs may be used to index both unexpected US presentations and omissions, yet this effect appears to be limited to relatively high levels of certainty in the control conditions.

The inclusion of online expectancy ratings in the third experiment allowed us to directly test the relationship between expectancy violations and outcome responses. We first tested if outcome responses differed between actually expected and actually unexpected outcomes, and found strong support for our main results, showing larger responses to unexpected outcomes. Furthermore, pupil responses to US presentation are not only able to distinguish between expected and unexpected outcomes, but also linearly relate to the size of the expectancy violation, showing stronger responses to more unexpected outcomes. This indicates that pupil responses to US presentations appear to track expectancy violations most reliably. Lastly, we tested whether outcome responding could predict an update of US expectancy from the present to the next trial of the same US probability. A change in US expectancy between two trials has previously been used to infer the experience of a prediction error (in an all-or-none manner) in a memory reconsolidation paradigm (Sevenster et al., 2013, 2014). In the current data, we did not find a relationship between outcome responses and changes in US expectancies on an individual level. This finding is in line with our other results showing that (for most measures) outcome response are not precise enough to reflect the exact size of the expectancy violation.

The crucial role of prediction errors in the acquisition and updating of fear memories has received increased attention over the past years, but a reliable physiological quantification of prediction errors is still lacking. We have found that both pupil dilation and SCRs at outcome can index unexpected compared to expected outcomes in aversive learning. To further assess whether these responses may indeed reflect prediction errors, we explored the relationship between outcome responses and changes in conditioned responding on a next trial. In contrast with our hypothesis where a positive prediction error (i.e., unexpected US presentation) would lead to a subsequent increase in conditioned responding and a negative prediction error (i.e., unexpected US omission) to a decrease in conditioned responding, we found that only pupil responses to US presentations reliably predicted an increase in conditioned responding. In Experiment 2, we did see a trend-wise significant relationship between pupil responses to US omissions and an increase in conditioned responding. While the absence of significant effects complicates the interpretation of these findings, the fact that US omission responses positively relate to conditioned responding directly contradicts our initial hypothesis. It should be noted that this hypothesis assumed that conditioned pupil responses reflect the associative
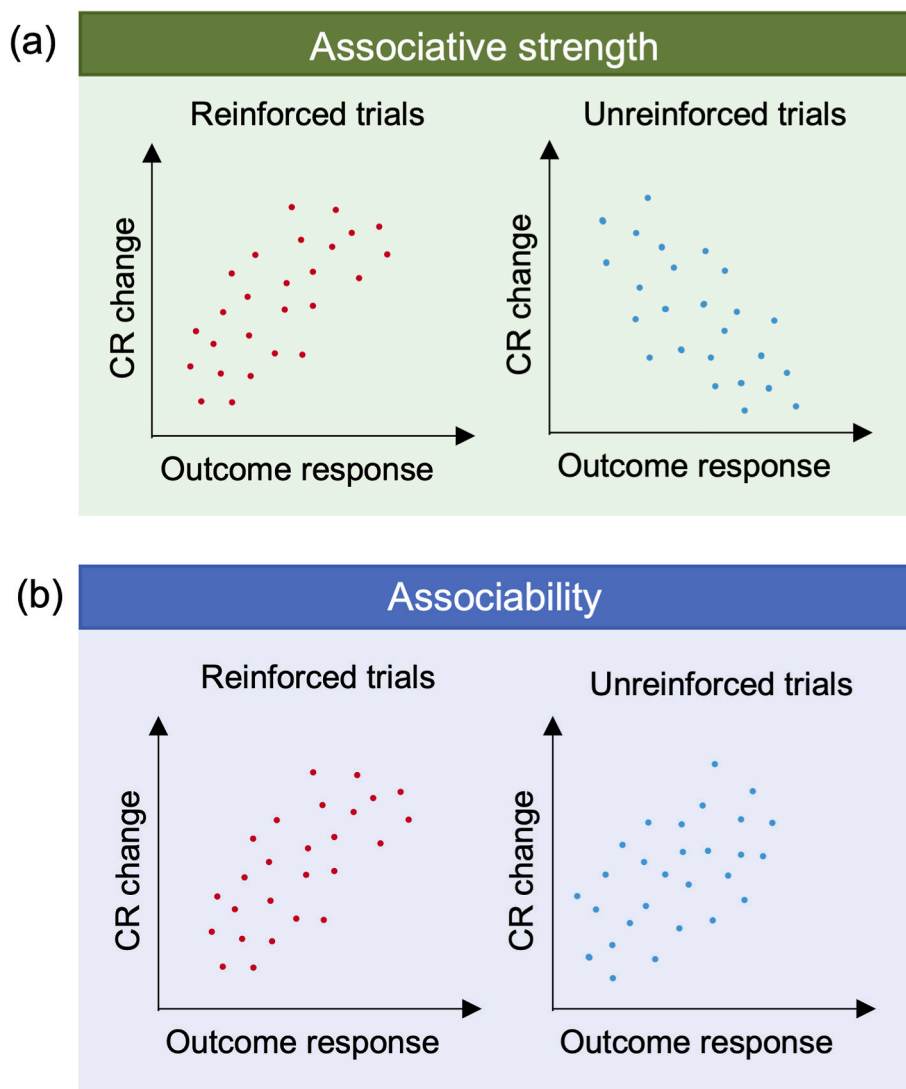

Fig. 6. Expected patterns of the relationship between outcome responses and a change in conditioned responding, dependent on the various parameters that may drive conditioned responding. a) The relationship between outcome responses and conditioned responding if the conditioned response reflects the associative strength (Rescorla & Wagner, 1972). b) The relationship between outcome responses and conditioned responding if the conditioned response reflects the associability (Pearce & Hall, 1980).

strength of the CS (Fig. 6a), whereas these responses may also reflect the associability of the CS. The associability of a stimulus is often understood in terms of the attention paid to the stimulus (Pearce & Hall, 1980), and is driven by unsigned prediction errors that occur during unexpected US presentations and unexpected US omissions (Fig. 6b). The pattern in our pupil data thus appears to best explained by outcome responses reflecting unsigned prediction errors, which increase the associability of the CS on a next trial. Indeed, pupil responses have previously been found to reflect the associability of a CS (Koenig et al., 2017; Ojala & Bach, 2020), and our conditioned pupil responses are mostly in line with this interpretation (i.e., they are largest for the CS_{50}). Nevertheless, the evidence for effects of omission responses on conditioned responding is rather weak, and this idea thus requires further investigation.

For SCRs we found no evidence that outcome responses predict any updating of the conditioned response on a next trial, which may be due to a lack of learning in the current design. While the experiments were designed to maximize prediction errors in a learning task, it was not necessary to update knowledge about the CS-US probabilities once this knowledge was acquired. The lack of contingency changes in the first and third experiments may have therefore attenuated the relationship between outcome responding and CS responding, especially for US omissions. The unreinforced trials of a 50% reinforced CS increase uncertainty about the occurrence of the US, but do not necessarily prompt safety learning. The second experiment included stronger contingency changes, and in line with our predictions, we observed a trend-wise significant interaction between outcome responses and reinforcement. While this interaction should be interpreted carefully, it may suggest that – if there is a stronger necessity for learning – outcome responses could predict an update in conditioned responding in line with prediction error-driven learning. Ultimately, while the multilevel analyses performed here can provide interesting insights into the relationship between outcome responses and conditioned responses, computational models should be fitted to the data to robustly relate observed responses to parameters of associative learning models. Further, an experimental task with a stronger focus on (re)learning such as an extinction paradigm could provide better insights into the relationships between outcome responses and subsequent learning.

In sum, our findings suggest that both SCR and pupil dilation can index (unsigned) prediction errors during outcome responding, which in the case of pupil responses may increase associability on a next trial (Pearce & Hall, 1980). Nonetheless, there are some important considerations to (the interpretation of) our findings. First, these findings were not entirely consistent across the experiments. One important addition to the third experiment compared to the first, is that participants were asked to provide US expectancy ratings for each trial. These ratings, intended as manipulation check, may have increased the formation of specific expectations about the outcome, and thereby strengthened our results. Moreover, the effects we found in the last two experiments appear strongly dependent on a high level of certainty about the control conditions and experimental instructions. Differences between expected and unexpected outcomes may thus not be strong enough to be observed under more ambiguous circumstances, such as in Experiment 1. In support of this, our analyses of the expectancy data in Experiment 3 show that outcome responses for SCR data do not linearly relate to the level of unexpectedness of the outcome, which would be expected from a clean prediction error signal. In a similar vein, Willems and Vervliet (2021) found that US omission responses only differed between fully expected omissions and unexpected omissions, but not between the various levels of unexpectedness (i.e., 75% probability versus 25% probability). Lastly, in all current experiments, data from multiple trials is pooled together to reduce noise. This means that the observed differences may be driven by some trials more than others, which is not in line with a prediction error signal. To investigate the role of prediction errors in fear and extinction learning, an index of prediction errors should be strong enough to measure subtle variations in prediction error

occurrence during a single trial, for which both SCR and pupil responses may be too noisy.

Another important consideration when using physiological measures to index prediction errors is the potential influence of CS responses on outcome responses. While pupil responses tend to be relatively fast with short response tails (Korn et al., 2017; Ojala & Bach, 2020), SCRs can last for more than 5 seconds (Boucsein et al., 2012), meaning that a CS-driven response may continue beyond CS offset and interfere with the outcome response. Manually scoring the SCR data allowed us to specifically identify new responses, marked by a clear onset, rather than unknowingly including ongoing responses. Nevertheless, while this procedure ensures the inclusion of distinct responses on downward or upward slopes, it does not completely exclude the possibility that the ongoing CS response affects the outcome response. The use of a toolbox like PsPM (Bach et al., 2013), which applies a general linear model to estimate the relative contributions of each event (CS onset, outcome) could potentially circumvent this problem. However, this requires jittering of the CS duration, which may in turn induce timing-related prediction errors. Importantly, this problem of potential influence of CS responses is inherent to most designs aiming to measure outcome responses following a preceding CS response. Here, we tried to the best of our ability to control for these influences, but the results should be interpreted in light of this limitation.

In conclusion, the current study shows that both pupil dilation (for US presentations only) and skin conductance can index unexpected outcomes compared to expected outcomes. Crucially, pupil responses to US presentations predict an increase in conditioned responding on a next trial, and the absence of an inverse relationship for omission responses may suggest the involvement of attentional processes. We found no evidence for a relationship between outcome responses and conditioned responding for SCRs, potentially because there is no strong requirement to update contingency knowledge in the current design. Furthermore, while we suggest that the inconsistencies between the experiments mostly arise from the inclusion of fewer trials and higher general uncertainty rather than from the non-existence of the effect, the null results that we observe in some of the experiments show that the effect of expectation on outcome responding is not robust in our current experimental design. The question remains whether outcome responses reflect prediction error magnitude with a precision that would be necessary for application at a single-trial level. Varying outcome probability and timing may provide further insights into how SCR and pupil responses are affected by outcome expectation, while for example extinction learning paradigms can elucidate whether outcome responses can also predict updating of the behavioral fear response. The development of a direct index of prediction errors will contribute to a more detailed understanding of how humans learn from unexpected events. We believe that physiological measurements of outcome responding can be a promising method to develop indices of unexpected outcomes under controlled circumstances, but may unfortunately not be precise enough for (sub)clinical application.

CRediT authorship contribution statement

Lotte E. Stemerding: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Vanessa A. van Ast:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Anna M.V. Gerlicher:** Methodology, Writing – review & editing. **Merel Kindt:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Declarations of competing interest

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2022.104164.

References

Bach, D. R., & Friston, K. J. (2012). No evidence for a negative prediction error signal in peripheral indicators of sympathetic arousal. *NeuroImage, 59*(2), 883–884. https://doi.org/10.1016/j.neuroimage.2011.08.091

Bach, D. R., Friston, K. J., & Dolan, R. J. (2013). An improved algorithm for model-based analysis of evoked skin conductance responses. *Biological Psychology, 94*(3), 490–497. https://doi.org/10.1016/j.biopsycho.2013.09.010

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Braem, S., Coenen, E., Bombeke, K., van Bochove, M. E., & Notebaert, W. (2015). Open your eyes for prediction errors. *Cognitive, Affective, & Behavioral Neuroscience, 15*(2), 374–380. https://doi.org/10.3758/s13415-014-0333-4

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience, 18*(4), 590–596. https://doi.org/10.1038/nn.3961

Colizoli, O., De Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. *Scientific Reports, 8*(1), 1–13. https://doi.org/10.1038/s41598-018-31985-3

Craske, M. G., Hermans, D., & Vervliet, B. (2018). State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*. https://doi.org/10.1098/rstb.2017.0025, 1742.

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Dunsmoor, J. E., Bandettini, P. A., & Knight, D. C. (2008). Neural correlates of unconditioned response diminution during Pavlovian conditioning. *NeuroImage, 40*(2), 811–817. https://doi.org/10.1016/j.neuroimage.2007.11.042

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Gazendam, F. J., Kamphuis, J. H., & Kindt, M. (2013). Deficient safety learning characterizes high trait anxious individuals. *Biological Psychology, 92*(2), 342–352. https://doi.org/10.1016/j.biopsycho.2012.11.006

Goodman, A. M., Harnett, N. G., & Knight, D. C. (2018). Pavlovian conditioned diminution of the neurobehavioral response to threat. *Neuroscience & Biobehavioral Reviews, 84*, 218–224. https://doi.org/10.1016/j.neubiorev.2017.11.021

Iordanova, M. D., Yau, J. O.-Y., McDannald, M. A., & Corbit, L. H. (2021). Neural substrates of appetitive and aversive prediction error. *Neuroscience & Biobehavioral Reviews, 123*, 337–351. https://doi.org/10.1016/j.neubiorev.2020.10.029

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron, 89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kindt, M. (2014). A behavioural neuroscience perspective on the aetiology and treatment of anxiety disorders. *Behaviour Research and Therapy, 62*, 24–36. https://doi.org/10.1016/j.brat.2014.08.012

Kloosterman, N. A., Lamme, V. A. F., Donner, T. H., van Loon, A. M., Bonneh, Y. S., & Meindertsma, T. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience, 41*(8), 1068–1078. https://doi.org/10.1111/ejn.12859

Knight, D. C., Lewis, E. P., & Wood, K. H. (2011). Conditioned diminution of the unconditioned skin conductance response. *Behavioral Neuroscience, 125*(4), 626–631. https://doi.org/10.1037/a0024324

Knight, D. C., Waters, N. S., King, M. K., & Bandettini, P. A. (2010). Learning-related diminution of unconditioned SCR and fMRI signal responses. *NeuroImage, 49*(1), 843–848. https://doi.org/10.1016/j.neuroimage.2009.07.012

Koenig, S., Uengoer, M., & Lachnit, H. (2017). Pupil dilation indicates the coding of past prediction errors: Evidence for attentional learning theory. *Psychophysiology, 1–12*.

Korn, C. W., Staib, M., Tzovara, A., Castegnetti, G., & Bach, D. R. (2017). A pupil size response model to assess fear learning. *Psychophysiology, 54*(3), 330–343. https://doi.org/10.1111/psyp.12801

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laurent, V., Balleine, B. W., & Westbrook, R. F. (2018). Motivational state controls the prediction error in Pavlovian appetitive-aversive interactions. *Neurobiology of Learning and Memory, 147*(September 2017), 18–25. https://doi.org/10.1016/j.nlm.2017.11.006

Lavín, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience, 7*, 1–8. https://doi.org/10.3389/fnbeh.2013.00218

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience, 15*(7), 1040–1046. https://doi.org/10.1038/nn.3130

Nieuwenhuis, R., te Grotenhuis, M., & Pelzer, B. (2012). Influence.ME: Tools for detecting influential data in mixed effects models. *R Journal, 4*(2), 38–47. https://doi.org/10.32614/rj-2012-011

Ojala, K. E., & Bach, D. R. (2020). Measuring learning in human classical threat conditioning: Translational, cognitive and methodological considerations. *Neuroscience & Biobehavioral Reviews, 114*, 96–112. https://doi.org/10.1016/j.neubiorev.2020.04.019

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America, 110*(38). https://doi.org/10.1073/pnas.1305373110

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Peterson, R., & Reiss, S. (1992). Anxiety sensitivty index manual. In *International diagnostic systems* (2nd ed.).

Preuschoff, K., Hart, B. M. 't, & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience, 5*, 1–12. https://doi.org/10.3389/fnins.2011.00115

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: Evidence from pupillometry and event-related fMRI. *NeuroImage, 37*(3), 1017–1031. https://doi.org/10.1016/j.neuroimage.2007.04.066

Sevenster, D., Beckers, T., & Kindt, M. (2012). Retrieval per se is not sufficient to trigger reconsolidation of human fear memory. *Neurobiology of Learning and Memory, 97*(3), 338–345. https://doi.org/10.1016/j.nlm.2012.01.009

Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction error governs pharmacologically induced amnesia for learned fear. *Science, 339*(6121), 830–833.

Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learning & Memory, 21*(11), 580–584. https://doi.org/10.1101/lm.035493.114

Sjouwerman, R., & Lonsdorf, T. B. (2019). Latency of skin conductance responses across stimulus modalities. *Psychophysiology, 56*(4), 1–11. https://doi.org/10.1111/psyp.13307

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience, 20*(10), 635–644. https://doi.org/10.1038/s41583-019-0180-y

Spielberger, C. D. (1983). *State-trait anxiety inventory for adults (STAI-AD)*. Palo Alto, CA: Consulting Psychologists Press.

Spoormaker, V. I., Andrade, K. C., Schröter, M. S., Sturm, A., Goya-Maldonado, R., Sämann, P. G., & Czisch, M. (2011). The neural correlates of negative prediction error signaling in human fear conditioning. *NeuroImage, 54*(3), 2250–2256. https://doi.org/10.1016/j.neuroimage.2010.09.042

Spoormaker, V. I., Blechert, J., Goya-Maldonado, R., Sämann, P. G., Wilhelm, F. H., & Czisch, M. (2012). Additional support for the existence of skin conductance responses at unconditioned stimulus omission. *NeuroImage, 63*(3), 1404–1407. https://doi.org/10.1016/j.neuroimage.2012.08.050

Sutton, R. S., & Barto, A. G. (1987). A temporal-difference model of classical conditioning. In *Proceedings of the ninth conference of the cognitive science society* (pp. 355–378). citeulike-article-id:4202914.

The MathWorks Inc. (2018). MATLAB. No. R2018b.

Tronson, N. C. (2020). Uncertainty versus prediction error in Pavlovian fear conditioning: Commentary on Walker et al. *European Journal of Neuroscience, 52*(5), 3485–3486. https://doi.org/10.1111/ejn.14578

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications, 8*, 1–11. https://doi.org/10.1038/ncomms14637

Van Slooten, J. C., Jahfari, S., Knapen, T., & Theeuwes, J. (2018). How pupil responses track value-based decision-making during and after reinforcement learning. *PLoS Computational Biology, 14*(11), 1–24. https://doi.org/10.1371/journal.pcbi.1006632

Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. *Nature Neuroscience, 16*(4), 1–16. https://doi.org/10.1038/nn.3345

Willems, A. L., & Vervliet, B. (2021). When nothing matters: Assessing markers of expectancy violation during omissions of threat. *Behaviour Research and Therapy, 136*, 1–10. https://doi.org/10.1016/j.brat.2020.103764