

In search of the behavioural effects of fear: a paradigm to assess conditioned suppression in humans

Gerlicher AMV^{1*}, Metselaar VN¹, Kindt M¹

¹Department for Clinical Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018WS Amsterdam, The Netherlands

*** Corresponding author**

Anna MV Gerlicher
Department of Clinical Psychology
Nieuwe Achtergracht 129B
1018WS Amsterdam
a.m.v.gerlicher@uva.nl
+31 20 5256 810

ABSTRACT

Conditioned fear can substantially reduce the likelihood that an individual will engage in reward-related behaviour - a phenomenon coined conditioned suppression. Despite the unmistakable relevance of conditioned suppression for excessive fears and their adverse consequences, the phenomenon has primarily been observed in animal models and is not yet well understood. Here, we aimed to develop a conditioned suppression paradigm that enables a robust quantification of the effect of Pavlovian aversive stimuli on subsequent reward-related behaviour in humans and assess its potential relation to physiological measures of fear. In phase 1, an instrumental response was incentivized with monetary rewards. In phase 2, one of two conditioned stimuli (CS+) was reinforced with an aversive unconditioned stimulus (US, i.e., electric stimulus). During aversive Pavlovian learning we assessed differential skin conductance (SCR) and fear potentiated startle responses (FPS). Lastly, we tested the effect of the aversively conditioned CS+ on the response rate of the instrumental response in a transfer phase. Despite strong aversive Pavlovian conditioning, as indicated by large effect sizes in differential SCR and FPS, we did not find any evidence for conditioned suppression: i.e., there was no significant reduction of instrumental responding in the presence of the CS+ compared to a new control stimulus. This lack of conditioned suppression is in line with previous studies that reported difficulties inducing conditioned suppression and points towards a general challenge in investigating conditioned suppression in humans. Implications and directions for future research on the highly relevant behavioural effects of fear and anxiety are discussed.

1. INTRODUCTION

To optimize our chances for survival and prepare for situation-appropriate action, fear and anxiety can interfere with ongoing goal-directed behaviour, such as going to work or engaging in social interactions. In face of actual threat, behavioural effects of fear are adaptive, but they can cause great harm when the actual threat is low and fear excessive. When goal-directed actions are repeatedly or constantly disrupted, even though the costs of interrupting the current action are high, an individual will experience less frequent rewards. The fear-induced reduction of rewarding activities could be a precursor for developing comorbid depression, or as stated in Hippocrates Epidemics, long-lasting fright may turn into melancholy (Hippocrates, 1868). In line with this thinking, the majority of comorbid mood and anxiety disorders are characterized by anxiety disorder symptoms preceding symptoms of depression (e.g., Cole et al. 1998; Kaufman and Charney 2000; Moitra, Herbert, and Forman 2008; Starr and Davila 2012; Wittchen et al. 2000). Recent years have seen a renewed interest in other, more direct behavioural consequences of conditioned fear, such as active and passive avoidance or escape (Dymond, 2019; Krypotos, Eftting, Kindt, & Beckers, 2015; LeDoux, Moscarello, Sears, & Campese, 2017; Pittig, Wong, Glück, & Boschet, 2020). The more indirect or subsequent effects of fear on reward-related behaviour are, however, largely understudied in humans. One reason for this apparent gap in the literature may be the lack of a suitable paradigm to examine this phenomenon in humans.

In animals it has been well established that a conditioned stimulus (CS) acquires some of the incentive motivational properties of the aversive stimulus with which it was paired during Pavlovian conditioning. These incentive motivational effects of a CS are expressed in multiple ways of which the most extensively studied are freezing (for reviews see e.g., Fanselow, 1994; Maren, 2001) and avoidance or escape behavior (for review see e.g., LeDoux et al., 2017). More recently, the Pavlovian-to-Instrumental transfer (PIT) paradigm has received more attention as it allows a differentiation between instrumental and Pavlovian learning and demonstrates how CSs can ignite behavioral responses other than those directly conditioned to the CS (LeDoux et al.,

2017). More specifically, the PIT or transfer paradigm has been developed to test the effects of conditioned fear on subsequent instrumental behaviour (Cartoni, Balleine, & Baldassarre, 2016; Holmes, Marchand, & Coutureau, 2010). It typically consists of three phases: (1) a Pavlovian phase, in which an initially neutral stimulus (CS) predicts the occurrence of an aversive stimulus (unconditioned stimulus, US), (2) an instrumental phase, in which a behaviour is reinforced by the omission of an aversive or the delivery of a rewarding stimulus, and (3) a transfer phase, in which the behavioural response can be exerted again in the presence of the Pavlovian CS. When confronted with the aversively conditioned CS, animals increase avoidance responses that lead to an omission of an anticipated aversive stimulus - a phenomenon termed conditioned facilitation (Rescorla & LoLordo, 1965; Solomon & Turner, 1962). Conditioned facilitation is similar to active avoidance as studied in avoidance paradigms, however, in conditioned facilitation the Pavlovian CS can also increase avoidance responses that were trained with a different US (e.g., loud noise) than the Pavlovian CS itself (e.g., electric stimulus; Hendersen, Patterson, & Jackson, 1980; LoLordo, 1967). Importantly, the presentation of the Pavlovian CS also reduces responding associated with a rewarding outcome, a phenomenon termed *conditioned suppression* (Estes & Skinner, 1941). Whereas Pavlovian-to-instrumental transfer paradigms have long been employed to study the effect of aversive learning on reward-related behaviour in animals, up-to-date a robust paradigm to investigate conditioned suppression in humans is lacking.

The small number of previous studies on conditioned suppression in humans were not consistently successful in showing a PIT effect (for review see Gerlicher & Kindt, 2020). Commonly, an 'aversive' US was employed that bears little ecological validity for the study of fear (e.g., bitter juice, small monetary loss, or a loss of points in a computer game), or the instrumental response was merely instructed, and/or simply reinforced with instructed rewards such as points in a computer game (J. A. Di Giusto, Di Giusto, & King, 1974; Hebart & Gläscher, 2015; Trick, Hogarth, & Duka, 2011). This makes comparisons to real-life reward behaviour acquired by actual experience and reinforced by appetitive stimuli difficult. In a recent study these two issues were

addressed by investigating the effect of a Pavlovian CS paired with an electric stimulus on an acquired instrumental response reinforced with a primary food reward (i.e., chocolate; Xia, Gurkina, & Bach, 2019). Though different from classic conditioned suppression paradigms in animals, a response conflict (Go/noGo task) was used to assess the motivational effects of fear conditioning on behaviour. Participants were asked to either withhold responding (No Go-conditions), or direct a coin towards ('approach') or away ('withdraw') from a target (Go-conditions) while a previously aversively conditioned CS (CS+) or a control stimulus that was never paired with the US (CS-) were presented in the background. In the presence of the CS+ compared to the CS-, a facilitation of withdrawal-responses was observed. However, no conditioned suppression of approach-responses was observed. Critically, participants only received the reward when the number of button presses resulted in a coin hitting a spatial target window. This may have limited the range of successful responses and potentially made response-rates less sensitive to more subtle effects of conditioned suppression. Furthermore, the paradigm did not include a baseline condition. As the CS- has been a reliable predictor of the absence of the US it can become an inhibitory or safety stimulus with its own effects on behaviour (Rescorla, 1969). The lack of a baseline condition can thus make it difficult to differentiate effects of an aversively trained CS+ from potential effects of the CS-. Here, we aimed to test a paradigm that more closely resembles PIT paradigms developed for animals and addressed potential caveats of previous PIT studies in humans. Hence, we did not employ any response conflict (Go/NoGo), allowed participants to respond freely and as often they wanted, and included a baseline condition to discern effects of the aversively conditioned Pavlovian CS+ from the effect of a potentially inhibitory CS-.

An outstanding question is how aversive Pavlovian learning relates to later effects of the CS on instrumental behavior. To answer this questions, reliable indices of aversive Pavlovian learning are necessary. A potential shortcoming of previous human PIT studies is thus that

aversive Pavlovian conditioning was hardly ever assessed with physiological measures of fear. In the studies that did assess physiological measures of fear, skin conductance responses (SCR) and pupil dilation were used (e.g., Garofalo & Robbins, 2017; Xia et al., 2019). SCR however may not reflect the expected aversive value of a Pavlovian CS, but rather its associability (Pearce & Hall, 1980), i.e., the degree to which the CS can become associated with the US (Li, Schiller, Schoenbaum, Phelps, & Daw, 2011; Seymour et al., 2005; Tzovara, Korn, & Bach, 2018; Zhang, Mano, Ganesh, Robbins, & Seymour, 2016). Similarly, some suggest that pupil dilation reflects uncertainty about the US instead of the expected value of the CS (Koenig, Uengoer, & Lachnit, 2018). Since neither SCR nor pupil dilation capture the expected aversive value of the CS (i.e, its associative strength) they may be sub-optimal indices of the incentive motivational properties of the aversive stimulus (US). This may explain the absence of any relationship between these physiological measures of fear and the behavioural effects of a Pavlovian CS. Fear potentiated startle responses (FPS) on the other hand have been shown to be resistant to outcome devaluation - a key feature of model-free value learning (Dayan & Berridge, 2014). That is, startle responses to an aversively conditioned CS remained high despite instructed extinction or the actual removal of the US electrode (Sevenster, Beckers, and Kindt 2012a, 2012b; but see Mertens et al. 2021). If the startle response to the CS is indeed more sensitive to reflect the expected value of a CS and the incentive motivational properties of the aversive stimulus (US), it may also be a predictor of the effect of a Pavlovian CS on behaviour. Here, we aimed to test this hypothesis by collecting both FPS and SCR and assessed whether FPS at the end of conditioning can predict how strongly an aversively trained Pavlovian CS interferes with instrumental reward behaviour.

Specifically, we employed a PIT paradigm comprising three phases (see Figure 1). In an instrumental learning phase (phase 1), participants were asked to learn the sequence in which three keyboard buttons needed to be pressed in order to receive a monetary reward (€0.50).

Correctly entered sequences were reinforced with a cash-sound and the presentation of a 50 Cent coin on the screen until each participant had earned €8.00. To ensure that the instrumental behaviour was sufficiently incentivized, the reward was paid out in cash directly after phase 1. In the subsequent Pavlovian learning phase (phase 2), participants were presented with two conditioned stimuli (CS), one of which (CS+) was reinforced with an uncomfortable electric stimulus whereas the other was never reinforced and served as a control stimulus (CS-). The success of Pavlovian aversive conditioning was assessed as differential (CS+>CS-) skin conductance (SCR) and fear potentiated startle response (FPS). In the third phase, participants could exert the instrumental response again while either the CS+, CS-, or a new control stimulus (i.e., that had not been seen before) were presented on the screen. In order to exclude that response suppression is caused by the actual delivery of the US instead of the acquired CS-US association, the transfer test was conducted in extinction, i.e., correct button presses were not reinforced by a monetary reward and CS+ presentations were not followed by the US anymore. We hypothesized that the presentation of the aversively conditioned CS+ compared to a new control stimulus would result in a reduction of instrumental responding (i.e., *conditioned suppression*). Furthermore, if the CS- acquired any safety properties during Pavlovian learning, its presentation might increase instrumental responding compared to the presentation of the new control stimulus (i.e., *conditioned facilitation*). Lastly, we were interested in whether the strength of Pavlovian conditioning could serve as a predictor of conditioned suppression. Specifically, we expected that greater differential fear potentiated startle, but not skin conductance responses, at the end of the Pavlovian learning phase would predict greater conditioned suppression in the transfer phase.

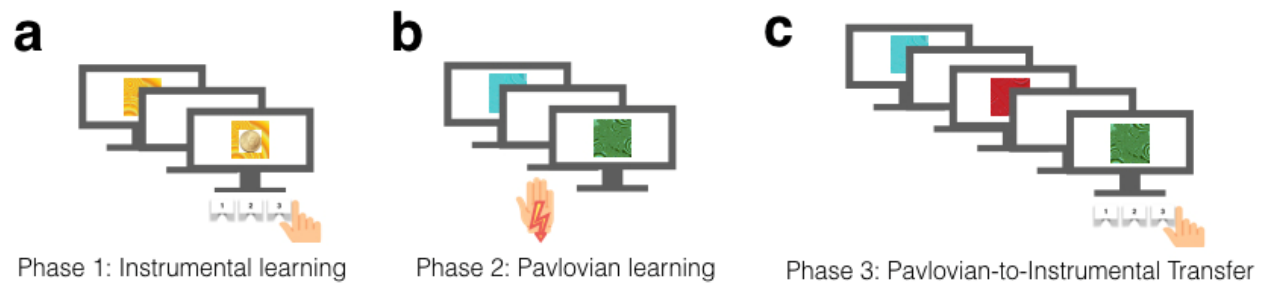


Figure 1 | Overview of the experimental design. The experiment consisted of three phases: **(a)** An instrumental learning phase in which participants learned which sequence of three button-presses was reinforced with a monetary reward of 50 Cent. The correct sequence was only reinforced when an image was presented on the screen. The earned amount was paid out directly after phase 1. **(b)** In the second phase, the Pavlovian learning phase, participants were presented with two new images, one of which was reinforced by an electric stimulus in 50% of the trials (CS+), the other one was never reinforced (CS-). **(c)** In the last phase, participants could perform the acquired response sequence again while either a new control stimulus, the aversively conditioned CS+ or the CS- were presented on screen. The Pavlovian-to-Instrumental test phase took place in extinction: neither the instrumental response nor the CS+ were reinforced.

2. METHODS AND MATERIALS

2.1 Participants

Based on an effect size of Cohen's $d=.40$ reported by a previous aversive transfer study in human participants (Xia et al., 2019), a power of $1-\beta=.80$, and an alpha error probability of $\alpha=.05$, we estimated a required sample size of $N=52$ using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). A pilot study in which we observed conditioned suppression with an effect size of $d_2=.78$ confirmed that this sample size would be sufficient to achieve a power of $1-\beta=.99$ (see Supplementary Figure 1). Before inclusion into the experiment, participants were screened for color-blindness. In total, $N=60$ participants were recruited for the experiment. Eight participants needed to be excluded ($N=1$ quit the experiment due to the aversiveness of the startle-probe, $N=2$ had participated in another fear conditioning study the same week, $N=5$ did not learn the correct order of button presses in the instrumental phase), leaving $N=52$ participants (mean 21.07 ± 2.59 years, range: 18-28 years; 37 female, 13 male, 1 other) for statistical analysis. The study was approved by the Ethics Review Board of the Department of Psychology of the

University of Amsterdam and all participants signed written informed consent. The study was conducted in accordance with the Declaration of Helsinki.

2.2 Questionnaires

For sample characterization, we collected questionnaire data of the trait-version of the Spielberger State-Trait-Anxiety Inventory (STAI-T; Spielberger et al., 1983) and the Anxiety Sensitivity Index (ASI-3; Taylor et al., 2007). Questionnaire data of one participant was missing due to a technical issue during data collection.

2.3 Stimuli

Four differently colored images (blue, yellow, green or red fractal, see Figure 1) were presented in the center of the screen and served either as instrumental stimulus, CS during Pavlovian learning, or new transfer stimulus during the PIT phase. Assignment of the four images to instrumental stimulus, CS+, CS- or transfer stimulus was randomized between participants.

2.4 Unconditioned stimuli (US)

An electric stimulus delivered to the wrist of the left hand via two 20 by 25 mm Ag/AgCl electrodes with fixed inter-electrodes mid-distances of 45 mm served as US. Conductive-gel was applied to the electric stimulus electrode (Signa Gel, Parker Laboratories Inc.). The electric stimulus itself consisted of a train of three square-wave pulses with a duration of 2 ms and an inter-pulse interval of 100 ms and 200 ms. The delivery of the US was controlled by a Digitimer DS7A constant current stimulator (Digitimer, Weybridge). Before the start of the experiment the intensity of the electric stimulus was calibrated individually to a level judged as 'maximally uncomfortable, but not yet painful' by the participant on a rating scale (i.e, 0='I do not feel anything', 5='medium uncomfortable', 10='already painful'). Subjective intensity ratings had a

mean \pm std of 9.14 \pm .74 (range: 9.00-9.90) and the actual intensity of the US ranged from 4.40-61.00 mA with a mean \pm std of 26.76 \pm 14.6 mA.

2.5 Fear potentiated startle response (FPS)

In order to elicit startle responses, a loud noise (40 ms; 104 dB) was presented binaurally via headphones (Model MD-4600; Compact Disk Digital Audio, Monacor). We recorded electromyographic (EMG) activity using 7 mm Ag/AgCl electrodes filled with conductive gel and positioned approximately 1 cm below the pupil and 1 cm below the lateral canthus, the outer corner of the eye (Fridlund & Cacioppo, 1986). A ground electrode was placed on an electrically neutral site on the forehead. The EMG signal was amplified and digitized at 1000 Hz. The signal was analyzed offline using Psycho-Physiological Modelling (PsPM 5.0.0; Bach et al. 2018) in Matlab 2020a (Mathworks®, Natick, Massachusetts, USA). In PsPM, the signal was rectified and band-pass filtered (cut-off: 50Hz and 470 Hz, 4th order Butterworth filter). Furthermore, a notch filter was applied to remove 50 Hz harmonics. The resulting signal was smoothed using a low-pass filter (cut-off: 53.05 Hz, 4th order Butterworth filter) and the data was down-sampled to 500 Hz (Khemka, Tzovara, Gerster, Quednow, & Bach, 2017). To estimate trial-by-trial startle responses we employed a single-trial general linear model (GLM) with one regressor for each startle-probe onset. The single-trial regressors were convolved with a canonical startle response function with a flexible response onset latency of 0-100 ms. Single-trial parameter estimates were Z-transformed across stimuli (CS+, CS-, NA; excluding habituation trials) and within each participant for statistical analysis.

2.6 Skin conductance response (SCR)

Electrodermal activity (EDA) was recorded from the middle phalanges of the index and middle finger of the left hand using two AG/AgCl Electrodes of 20 by 16 mm. The signal was recorded using a sine-shaped excitation voltage (5V) of 50 Hz derived from the mains frequency and was

digitized at 1000 Hz through a 16-bit AD-converter. The EDA signal was analyzed using PsPM 5.0.0. Specifically, we employed a single-trial GLM (Bach, Flandin, Friston, & Dolan, 2009, 2010) with one regressor for each CS onset and one regressor for each US delivery and a canonical SCR function with time-derivative and fixed response latency. For statistical analysis, single-trial parameter estimates were Z-transformed across stimuli (CS+, CS-) and within each participant.

2.7 Procedure

2.7.1 Instrumental phase

Upon arrival, participants filled out informed consent and questionnaires. Subsequently, the intensity of the US was calibrated. Before the start of the experiment, participants were informed that the experiment would consist of three phases and that they would be presented with a steady background noise throughout all three phases to shield them from environmental noise. Participants were then verbally instructed that they could earn money in the first phase by pressing three buttons in a specific order. The buttons (1, 2, 3 on the number pad) were marked with red dots on the keyboard. Whenever participants would press the buttons in the right order, they would be rewarded with 50 Cents, signaled by the presentation of a 50 Cent coin on the screen and a cash sound. The actual money earned throughout the phase would be paid out directly after the phase. Participants were instructed that they could only press the buttons to earn money when the fixation cross was not on the screen. They were told that they could earn as much money as they wanted by pressing the buttons in the correct order as often as possible, but that not every correct order would yield a reward. The first phase started with a written repetition of the instructions. During the instrumental phase, either one of the four images ('instrumental stimulus') or the fixation cross were presented in the center of the screen. The instrumental stimulus was presented for 8000 ms. During inter-trial intervals the fixation cross was presented in the center of the screen. The duration of ITIs was randomized between 15-20 seconds, with a

mean of 17.5 seconds. As in previous research (Weber et al., 2016), a variable-ratio schedule was faded in. That is, in the beginning a fixed-ratio reinforcement schedule was used in order to facilitate learning (i.e., 1, 1, 1, 1, 2, 4, 8, 16), subsequently, on average every 15th (range: 5-25) correct order was reinforced with 50 Cent. The first phase ended as soon as the participant had earned 8 €, which was on average after 32.06 ± 12.96 trials (range: 13-83). After the end of the first phase, the experimenter entered the room, paid out the reward directly to the participant and asked the participant to report the order of button presses that yielded the reward. Participants who could not report the sequence correctly were excluded from the experiment.

2.7.2 Pavlovian phase

Before the start of the Pavlovian phase, participants were instructed that they would see two different images and would occasionally receive an electric stimulus. Their task was to pay attention to any relationship between the images and the electric stimulus. Furthermore, they were instructed that they may hear loud noises during this phase. The Pavlovian phase started with the presentation of 10 NA trials for startle response habituation. Subsequently, 10 CS+, 10 CS-, 10 noise alone (NA) trials were presented. Trial order was randomized in such a way that no more than two trials of the same type were presented after each other. Each CS was presented for 8000 ms in total. The startle probe (104 dB, 40 ms) was delivered 7150 ms after stimulus onset and the train of three USs started 500 ms later in case of reinforced CS+. The CS+ was reinforced in 50% of trials. Reinforcement was randomized in such a way that not more than two succeeding CS+ presentations could be reinforced or unreinforced. During ITIs a fixation cross was presented in the center of the screen. The duration of the ITI was randomized between 15-20 seconds with a mean of 17.5 seconds.

2.7.3 PIT phase

Before the start of the transfer phase, participants were instructed that they could again earn money by pressing the buttons in the same order as in phase 1. During the transfer phase the CS+, CS-, and a new stimulus were presented for 6 trials each. Each trial lasted 8000 ms. During ITIs a fixation cross was presented in the center of the screen and the duration of the ITI was randomized between 15-20 seconds with a mean of 17.5 seconds. In contrast to the instrumental and Pavlovian learning phase, the transfer phase was conducted in extinction. That is, participants did not receive any monetary reward and no further USs were administered.

2.8 Statistical Analysis

Statistical analyses were conducted using RStudio (v1.1456, RStudio Team, 2016). To test whether participants acquired the correct order of button presses in the instrumental phase, we compared the number of correct responses averaged across the first two trials to the last two trials of the instrumental phase using a non-parametric Wilcoxon signed rank test. The success of Pavlovian learning was assessed by comparing conditioned responses (i.e., FPS, SCR) averaged across the first and last two trials of the Pavlovian phase using repeated-measures ANOVA with stimulus (CS+, CS-) and trial (first 2, last 2) as within-subject factors and Type-III sum of squares (ez-package v4.4.0; Lawrence, 2016). In pilot data with N=5 participants, we had observed that conditioned suppression of the number of correct responses was strongest on the first trial of the PIT phase (see Supplementary Figure 1). For this reason, we tested specifically whether conditioned suppression occurred on the first trial of the PIT phase by comparing the number of correct responses during the new control stimulus to the number of correct responses during CS+ using a paired sample t-test. Results were considered significant when $p < .05$ (two-sided tests).

3. RESULTS

The present sample had a mean±SD trait anxiety (STAI-T, Spielberger, Gorsuch, and Lushene 1970) score of 43.31±8.34 (range: 27-62) and a mean±SD anxiety sensitivity score (ASI-3; Taylor et al. 2007) of 17.82±11.96 (range: 2-59).

3.1 Phase 1 - Instrumental phase

All participants acquired the correct order of button presses as indicated by a significant increase of correct responses from the first to the last two trials of the instrumental phase (52 of 52 participants rank last 2 trials > first 2 trials; Wilcoxon Signed Rank Test, $Z=-6.27$, $p<.001$; see Figure 2).

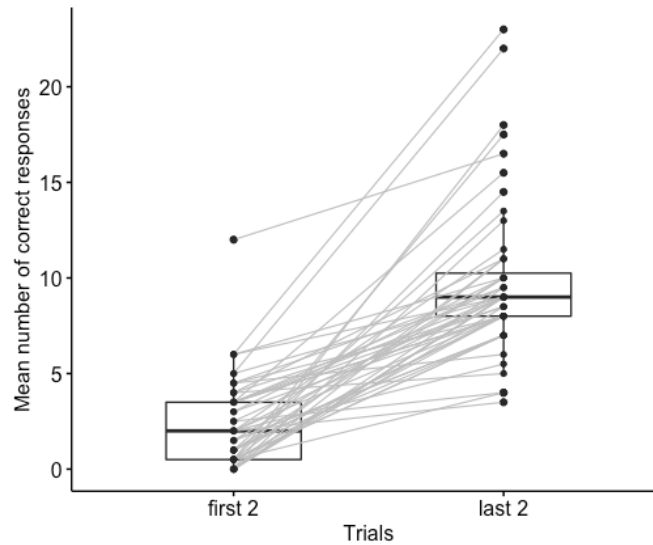


Figure 2 | Phase 1 - Instrumental learning. As to be expected after the exclusion of participants who did not learn the correct order, all participants showed a higher number of correct responses in the last two compared to the first two trials of the instrumental learning phase indicating a successful acquisition of the instrumental response.

3.2 Phase 2 - Pavlovian phase

Aversive learning was successful as indicated by a significant increase of differential (CS+>CS-) startle responses from the beginning to the end of the Pavlovian phase (stimulus: $F_{1,51}=30.89$, $p<.001$, $\eta^2_p= .38$; trial: $F_{1,51}=40.00$, $p<.001$, $\eta^2_p= .44$; stimulus* trial: $F_{1,51}=10.70$, $p=.002$, $\eta^2_p= .17$; Figure 3). In detail, there was no significant difference between FPS to CS+ and CS- during the

first two trials ($t_{51} = 1.94$, $p = .06$), but during the last two trials of conditioning ($t_{51} = 7.14$, $p < .001$, $d = .99$). Skin conductance responses also reflected successful acquisition of CS-US contingencies over the course of the Pavlovian learning phase (stimulus: $F_{1,51} = 130.63$, $p < .001$, $\eta^2_p = .72$; trial: $F_{1,51} = 23.03$, $p < .001$, $\eta^2_p = .31$; stimulus*trial: $F_{1,51} = .00$, $p = .99$, Figure 3). However, SCRs were significantly greater to CS+ than CS- during the first two trials already ($t_{51} = 7.37$, $p < .001$, $d = 1.02$) indicating that learning was reflected in SCR as soon as the participants had experienced the first CS+US and CS-noUS pairing. This stimulus effect was sustained up to the last two trials of conditioning ($t_{51} = 8.47$, $p < .001$, $d = 1.31$). Note, for both FPS and SCR effect sizes reflecting aversive learning were large (FPS: $d = .99$, SCR: $d = 1.31$).

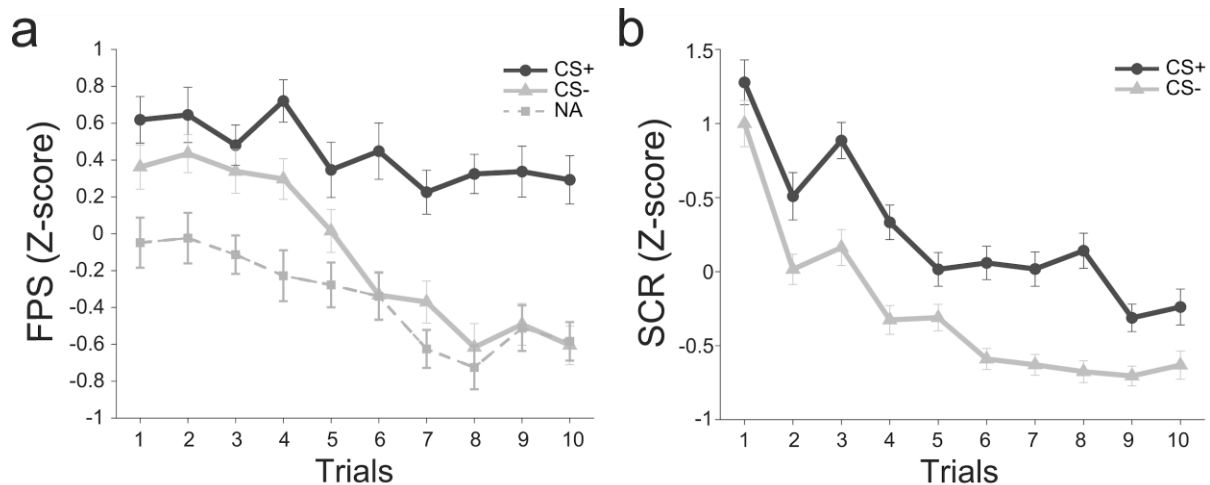


Figure 3 | Phase 2 - Pavlovian learning. Pavlovian learning was successful as indicated by (a) a significant increase in differential (CS+>CS-) fear potentiated startle responses from the beginning (first 2 trials) to the end (last 2 trials) of Pavlovian Learning, and by (b) significantly greater skin conductance responses to the CS+ than the CS- both at the beginning (first 2 trials) and the end (last 2 trials) of the Pavlovian phase. Error bars depict standard error of the mean (s.e.m.).

3.3 Phase 3 - Pavlovian-to-Instrumental transfer (PIT)

In the transfer phase, we compared the number of correct responses in the presence of the CS+ to a new baseline stimulus. In contrast to our hypothesis, we did not observe any significant difference between the number of correct responses during the CS+ compared to the new stimulus ($Z = -.74$, $p = .46$, see Figure 4). That is, the presence of the CS+ did not induce any

suppression of instrumental responding in the our experiment. There was also no significant difference between the number of correct responses during the CS- compared to the presentation of the new stimulus ($Z=-.71$, $p=.48$). This result suggests that the presentation of the CS- as a potential safety stimulus did also not evoke any conditioned facilitation. Furthermore, we assessed whether conditioned suppression may have been stronger in participants with high trait anxiety or high anxiety sensitivity. Even though there was indeed a positive relationship between trait anxiety and conditioned suppression, i.e. the number of correct responses in the presence of the new compared to the CS+; this relationship as well as the relationship between anxiety sensitivity and conditioned suppression were not significant (STAI-T: $\rho=.20$, $p=.16$; ASI: $\rho=.02$, $p=.90$). In additional exploratory analyses, we asked whether conditioned suppression was expressed as slower responding. Time between succeeding button presses did, however, not differ significantly between the CS+ (mean: 246.17 ± 77.08 ms, $t_{51}= 1.04$, $p=.30$) or the CS- (mean: 239.17 ± 72.53 ms, $t_{51}=-.21$, $p=.83$) compared to the new stimulus (mean: 240.55 ± 77.30 ms), even when looking at the first 1000 ms of the first trial only (see **Supplementary Figure 2**). In other words, despite strong aversive Pavlovian learning, neither the number of correct responses nor the time between succeeding button presses provided any evidence for conditioned suppression in the present study.

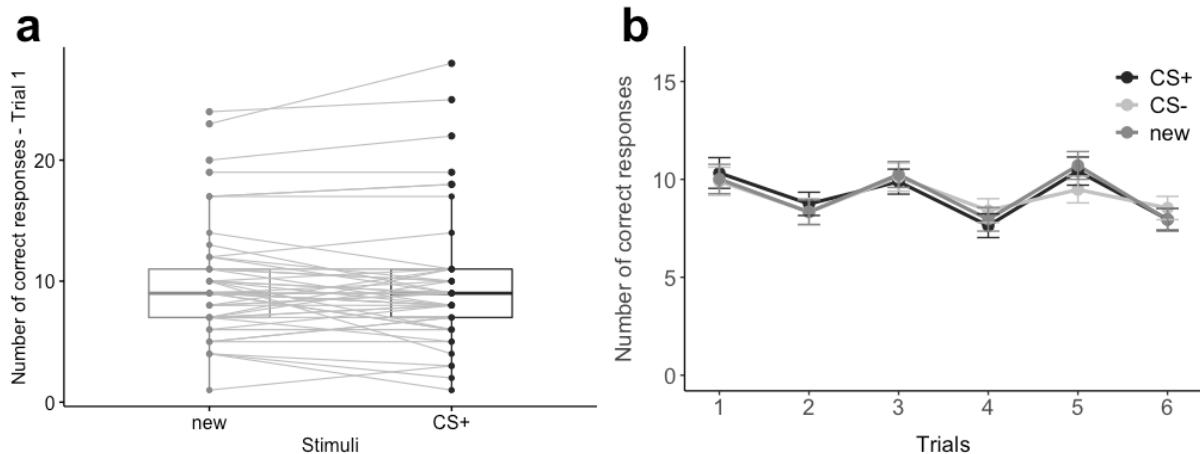


Figure 4 | Phase 3 - Number of correct responses during Pavlovian-to-Instrumental transfer. (a) In contrast to our hypothesis, there was no suppression of the number of correct responses in the presence of the CS+ compared to the new stimulus on the first trial of the PIT phase. (b) Number of correct responses across the six trials of the PIT phase did not differ significantly between the CS+, CS- and new stimulus (error bars depict standard error of the mean, s.e.m.).

We had furthermore hypothesized that the strength of Pavlovian learning assessed by FPS would predict the amount of conditioned suppression in the PIT phase. However, the relationship between differential (CS+>CS-) FPS at the end of conditioning and conditioned suppression as assessed by the number of correct responses in the presence of CS- compared to the CS+ was not significant (number correct responses: $\rho=.02$, $p=.88$). The result did not change when assessing conditioned suppression as time between succeeding responses ($r=-.02$, $p=.90$). In line with previous reports (Xia et al., 2019), there was also no significant relationship between differential SCR at the end of conditioning and either measure of conditioned suppression (number correct responses: $\rho= .04$, $p=.77$; time between responses: $r= .03$, $p=.84$).

4. Discussion

In the present study we aimed to develop a Pavlovian-to-Instrumental transfer paradigm to assess conditioned suppression after aversive conditioning in human participants. In contrast to previous studies, we used a primary aversive reinforcer during Pavlovian learning (i.e., an electric stimulus) and ensured that participants learned the instrumental response that resulted in a small monetary reward (secondary reinforcer) by trial and error. Furthermore, we employed a control stimulus in the transfer test in order to be able to dissociate effects from the aversively conditioned CS+ from potential effects of the CS-. In the instrumental phase of the experiment, the majority of recruited participants (52 out of 57) acquired the correct instrumental response within 32 trials. Participants who did not acquire the response were excluded from the experiment. Subsequent aversive Pavlovian conditioning as assessed by SCR and FPS was successful as reflected in large effect sizes. However, despite successful instrumental and Pavlovian learning, the presentation of the

aversively conditioned CS+ did not affect instrumental responding in the transfer phase. Instead participants showed a comparable number of correct responses during the CS+ and the new control stimulus. There was also no effect of the CS- on the response rate and we did not observe any relationship between individual differences in trait anxiety, anxiety sensitivity or measures of Pavlovian conditioning with conditioned suppression. Exploratory analyses of the time between succeeding responses in the first trial or even only the first 1000 ms of the first trial did not change the results. This indicates that the lack of a suppression effect cannot be explained by subtle suppression effects that extinguish quickly and therefore would only be detectable in the very beginning of the transfer phase.

Given the strong effect sizes in both measures of aversive Pavlovian conditioning (i.e., SCR and FPS) we can rule out that conditioning was not successful. It is, however, conceivable that aversive conditioning in human participants is not as impactful as in animals and employing a stronger US intensity would be necessary to induce conditioned suppression in the present human paradigm. This notion is indeed corroborated by animal research where conditioned suppression has been shown to be a function of US intensity, with stronger USs inducing stronger suppression (Annau & Kamin, 1961). Simply increasing the intensity of the here-employed electric stimulus would be unethical, but a multi-modal aversive stimulus as US could be an option to further enhance the motivational effects of aversive conditioning. A multimodal US could be composed of an image and a corresponding sound (e.g., sound of snapping celery - image of badly broken leg) and thereby emulate a real-life, multi-modal experience that amplifies the impact of a stimulus (Vries, Grasman, Kindt, & Ast, 2021). These stimuli do not pose a direct threat to the participant, but have been shown to be perceived as highly aversive and evoke a very strong physiological defensive responses (Vries et al., 2021). In the present study, we merely assessed whether a Pavlovian CS would elicit conditioned suppression. It bears mentioning that Pavlovian stimuli also facilitate avoidance responses after aversive conditioning – a finding that has been reported more consistently than conditioned suppression (e.g., Garofalo & Robbins, 2017; Xia et

al., 2019). As an additional manipulation check for aversive conditioning, the paradigm could be extended with conditions to assess the facilitation of avoidance responses. However, it has been shown that the same Pavlovian CS that reliably induces conditioned facilitation does not necessarily elicit conditioned suppression in the same participant (Xia et al., 2019), suggesting that the factors that modulate the induction of conditioned suppression differ from those that induce conditioned facilitation (i.e., avoidance behaviour).

Another critical factor determining the strength of conditioned suppression in animals is the motivational drive to perform the instrumental response (Millenson & de Villiers, 1972). It is therefore likely that also conditioned suppression in humans is substantially modulated by the value of the instrumental reward, with instrumental responses reinforced by outcomes with relatively low reward value being more easily suppressed than those reinforced with outcomes with high reward value. In line with this idea, studies employing instrumental reinforcers such as points in a computer game or art slides have previously reported to observe conditioned suppression (E. L. Di Giusto & Bond, 1978; J. A. Di Giusto et al., 1974; Punch, King, & Matyas, 1976). However, when the incentive to perform the instrumental action was increased by employing primary appetitive reinforcers such as chocolate (Xia et al., 2019) or secondary appetitive reinforcers such as money (present study, Hebart et al., 2015), studies failed to find conditioned suppression in humans. Future research should address to what extent conditioned suppression after human aversive conditioning is modulated by (a) the aversiveness of the Pavlovian US and (b) the reward value of the instrumental response, or their interaction.

A paradigm to assess conditioned suppression in humans could foster research into the causes of the detrimental behavioural consequences of fear, their neurobiological mechanisms and ways to overcome them. As an example, further exploring individual differences (e.g., reward sensitivity) as well as general factors (e.g., reward value) in modulating, amplifying and reducing conditioned suppression of reward behaviour could bring about new treatment approaches. A robust paradigm of conditioned suppression could in itself also be useful in translating insights

from basic to clinical science, for instance by directly investigating the behavioural effect of new interventions instead of solely using physiological read-outs. So far, the attempts of developing a paradigm of fear-induced conditioned suppression in humans have been unsuccessful, but additional research to further explore this phenomenon may be worthwhile.

Summarizing, the present study aimed to establish a Pavlovian-to-instrumental transfer paradigm to assess conditioned suppression after aversive Pavlovian conditioning in humans. Despite strong Pavlovian conditioning we did not observe an effect of the aversively trained CS on instrumental reward behaviour. Future studies may explore whether reducing the incentive to perform the instrumental response and/or increasing the intensity of the Pavlovian US may facilitate the induction of conditioned suppression in humans.

REFERENCES

- Annau, Z., & Kamin, L. J. (1961). The conditioned emotional response as a function of intensity of the US. *Journal of Comparative and Physiological Psychology*, *54*, 428–432. doi: 10.1037/h0042199
- Bach, D. R., Castegnetti, G., Korn, C. W., Gerster, S., Melinscak, F., & Moser, T. (2018). Psychophysiological modeling: Current state and future directions. *Psychophysiology*, *55*(11), e13214. doi: 10.1111/psyp.13209
- Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2009). Time-series analysis for rapid event-related skin conductance responses. *Journal of Neuroscience Methods*, *184*(2), 224–234. doi: 10.1016/j.jneumeth.2009.08.005
- Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2010). Modelling event-related skin conductance responses. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *75*(3), 349–356. doi: 10.1016/j.ijpsycho.2010.01.005
- Cartoni, E., Balleine, B., & Baldassarre, G. (2016). Appetitive Pavlovian-instrumental Transfer: A review. *Neuroscience and Biobehavioral Reviews*, *71*, 829–848. doi: 10.1016/j.neubiorev.2016.09.020
- Cole, D. A., Peeke, L. G., Martin, J. M., Truglio, R., & Seroczynski, A. D. (1998). A longitudinal look at the relation between depression and anxiety in children and adolescents. *Journal of Consulting and Clinical Psychology*, *66*(3), 451–460. doi: 10.1037/0022-006X.66.3.451
- Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 473–492. doi: 10.3758/s13415-014-0277-8
- Di Giusto, E. L., & Bond, N. W. (1978). One-trial conditioned suppression: Effects of instructions on extinction. *American Journal of Psychology*, *91*(2), 313–319.
- Di Giusto, J. A., Di Giusto, E. L., & King, M. G. (1974). Heart rate and muscle tension correlates of conditioned suppression in humans. *Journal of Experimental Psychology*, *103*(3), 515–521. doi: 10.1037/h0037207
- Dymond, S. (2019). Overcoming avoidance in anxiety disorders: The contributions of Pavlovian and operant avoidance extinction methods. *Neuroscience and Biobehavioral Reviews*, *98*, 61–70. doi: 10.1016/j.neubiorev.2019.01.007

- Estes, W. K., & Skinner, B. F. (1941). Some quantitative properties of anxiety. *Journal of Experimental Psychology*, 29(5), 390–400.
- Fanselow, M. S. (1994). Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin & Review*, 1(4), 429–438. doi: 10.3758/BF03210947
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, 23(5), 567–589.
- Garofalo, S., & Robbins, T. W. (2017). Triggering Avoidance: Dissociable Influences of Aversive Pavlovian Conditioned Stimuli on Human Instrumental Behavior. *Frontiers in Behavioral Neuroscience*, 11, 63. doi: 10.3389/fnbeh.2017.00063
- Gerlicher, A., & Kindt, M. (2020, September 15). *A review on aversive Pavlovian-to-Instrumental transfer in humans*. PsyArXiv. doi: 10.31234/osf.io/dyz29
- Hebart, M. N., & Gläscher, J. (2015). Serotonin and dopamine differentially affect appetitive and aversive general Pavlovian-to-instrumental transfer. *Psychopharmacology*, 232(2), 437–451. doi: 10.1007/s00213-014-3682-3
- Henderson, R. W., Patterson, J. M., & Jackson, R. L. (1980). Acquisition and retention of control of instrumental behavior by a cue-signaling airblast: How specific are conditioned anticipations? *Learning and Motivation*, 11(4), 407–426. doi: 10.1016/0023-9690(80)90026-0
- Hippocrates. (1868). *The Genuine Works of Hippocrates*. (C. Darwin Adams, Trans.). New York, NY, US.
- Holmes, N. M., Marchand, A. R., & Coutureau, E. (2010). Pavlovian to instrumental transfer: A neurobehavioural perspective. *Neuroscience and Biobehavioral Reviews*, 34(8), 1277–1295. doi: 10.1016/j.neubiorev.2010.03.007
- Kaufman, J., & Charney, D. (2000). Comorbidity of mood and anxiety disorders. *Depression and Anxiety*, 12(S1), 69–76. doi: [https://doi.org/10.1002/1520-6394\(2000\)12:1+<69::AID-DA9>3.0.CO;2-K](https://doi.org/10.1002/1520-6394(2000)12:1+<69::AID-DA9>3.0.CO;2-K)

- Khemka, S., Tzovara, A., Gerster, S., Quednow, B. B., & Bach, D. R. (2017). Modeling startle eyeblink electromyogram to assess fear learning. *Psychophysiology*, *54*(2), 204–214. doi: 10.1111/psyp.12775
- Koenig, S., Uengoer, M., & Lachnit, H. (2018). Pupil dilation indicates the coding of past prediction errors: Evidence for attentional learning theory. *Psychophysiology*, *55*(4). doi: 10.1111/psyp.13020
- Kryptos, A.-M., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: A review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience*, *9*. doi: 10.3389/fnbeh.2015.00189
- LeDoux, J. E., Moscarello, J., Sears, R., & Campese, V. (2017). The birth, death and resurrection of avoidance: A reconceptualization of a troubled paradigm. *Molecular Psychiatry*, *22*(1), 24–36. doi: 10.1038/mp.2016.166
- Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*, *14*(10), 1250–1252. doi: 10.1038/nn.2904
- LoLordo, V. M. (1967). Similarity of conditioned fear responses based upon different aversive events. *Journal of Comparative and Physiological Psychology*, *64*(1), 154–158.
- Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annual Review of Neuroscience*, *24*, 897–931. doi: 10.1146/annurev.neuro.24.1.897
- Mertens, G., Boddez, Y., Kryptos, A.-M., & Engelhard, I. M. (2021). Human fear conditioning is moderated by stimulus contingency instructions. *BIOLOGICAL PSYCHOLOGY*, *158*. doi: 10.1016/j.biopsycho.2020.107994
- Millenson, J. R., & de Villiers, P. A. (1972). Motivational properties of conditioned suppression. *Learning and Motivation*, *3*(2), 125–137. doi: 10.1016/0023-9690(72)90033-1
- Moitra, E., Herbert, J. D., & Forman, E. M. (2008). Behavioral avoidance mediates the relationship between anxiety and depressive symptoms among social anxiety disorder patients. *Journal of Anxiety Disorders*, *22*(7), 1205–1213. doi: 10.1016/j.janxdis.2008.01.002
- 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.

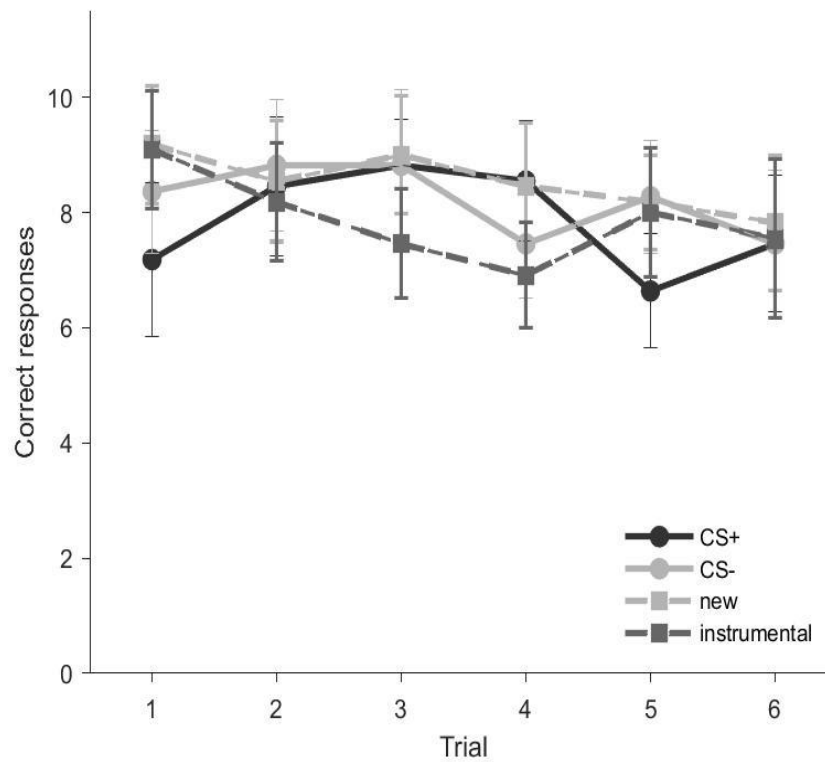
- Pittig, A., Wong, A. H. K., Glück, V. M., & Boschet, J. M. (2020). Avoidance and its bi-directional relationship with conditioned fear: Mechanisms, moderators, and clinical implications. *Behaviour Research and Therapy*, *126*, 103550. doi: 10.1016/j.brat.2020.103550
- Punch, J. C., King, M. G., & Matyas, T. A. (1976). ECG T-wave amplitude, muscle tension, and heart rate concomitants of conditioned suppression. *Physiological Psychology*, *4*(3), 294–302. doi: 10.3758/BF03332874
- Rescorla, R. A., & LoLordo, V. M. (1965). Inhibition of avoidance behavior. *Journal of Comparative and Physiological Psychology*, *59*(3), 406–412.
- Sevenster, D., Beckers, T., & Kindt, M. (2012a). Instructed extinction differentially affects the emotional and cognitive expression of associative fear memory. *Psychophysiology*, *49*(10), 1426–1435. doi: 10.1111/j.1469-8986.2012.01450.x
- Sevenster, D., Beckers, T., & Kindt, M. (2012b). Retrieval per se is not sufficient to trigger reconsolidation of human fear memory. *Neurobiology of Learning and Memory*, *97*(3), 338–345. doi: 10.1016/j.nlm.2012.01.009
- Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neuroscience*, *8*(9), 1234–1240. doi: 10.1038/nn1527
- Solomon, R. L., & Turner, L. (1962). Discriminative classical conditioning in dogs paralyzed by curare can later control discriminative avoidance responses in the normal state. *Psychological Review*, *69*(3), 202–219.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). Manual for the State-Trait Anxiety Inventory. *Consulting Psychologists Press*.
- Starr, L. R., & Davila, J. (2012). Temporal patterns of anxious and depressed mood in generalized anxiety disorder: A daily diary study. *Behaviour Research and Therapy*, *50*(2), 131–141. doi: 10.1016/j.brat.2011.11.005
- Taylor, S., Zvolensky, M. J., Cox, B. J., Deacon, B., Heimberg, R. G., Ledley, D. R., ... Cardenas, S. J. (2007). Robust dimensions of anxiety sensitivity: Development and initial validation of the Anxiety Sensitivity Index-3. *Psychological Assessment*, *19*(2), 176–188. doi: 10.1037/1040-3590.19.2.176

- Trick, L., Hogarth, L., & Duka, T. (2011). Prediction and uncertainty in human Pavlovian to instrumental transfer. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(3), 757–765. doi: 10.1037/a0022310
- Tzovara, A., Korn, C. W., & Bach, D. R. (2018). Human Pavlovian fear conditioning conforms to probabilistic learning. *PLOS Computational Biology*, *14*(8), e1006243. doi: 10.1371/journal.pcbi.1006243
- Vries, O. de, Grasman, R., Kindt, M., & Ast, V. van. (2021, April 6). *Threat learning impairs subsequent memory recombination with past episodes*. PsyArXiv. doi: 10.31234/osf.io/f7rww
- Wittchen, H.-U., Kessler, R. C., Pfister, H., Höfler, M., & Lieb, R. (2000). Why do people with anxiety disorders become depressed? A prospective-longitudinal community study. *Acta Psychiatrica Scandinavica*, *102*(s406), 14–23. doi: 10.1111/j.0065-1591.2000.acp29-03.x
- Xia, Y., Gurkina, A., & Bach, D. R. (2019). Pavlovian-to-instrumental transfer after human threat conditioning. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *26*(5), 167–175. doi: 10.1101/lm.049338.119
- Zhang, S., Mano, H., Ganesh, G., Robbins, T., & Seymour, B. (2016). Dissociable Learning Processes Underlie Human Pain Conditioning. *Current Biology: CB*, *26*(1), 52–58. doi: 10.1016/j.cub.2015.10.066

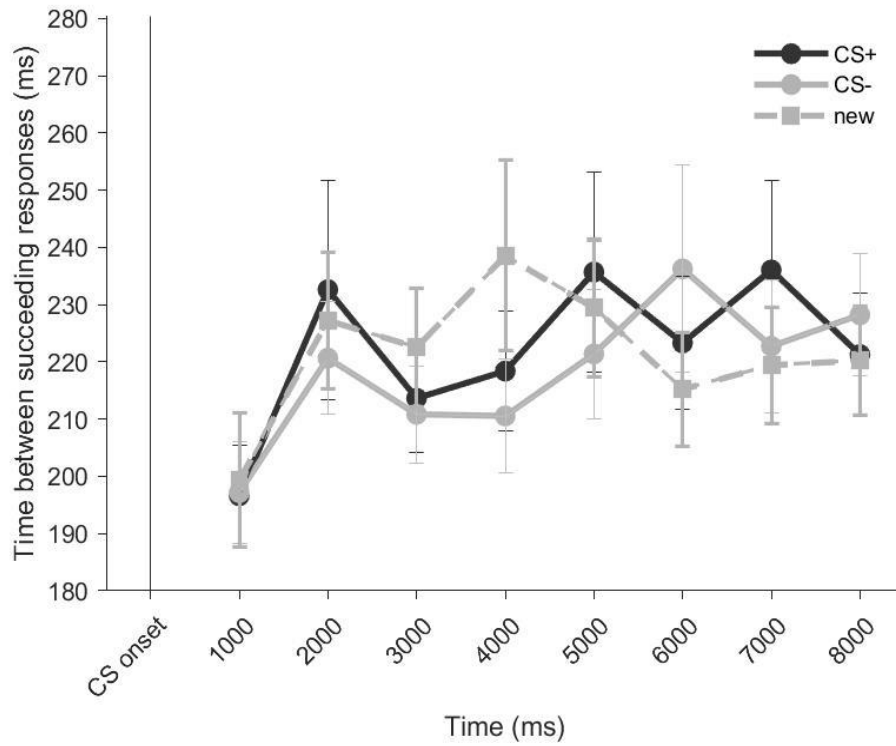
Author Notes

We thank Bert Molenkamp for technical support. The present work was supported by a European Research Council (ERC) Advanced Grant 743163 to M.K. Data that support the findings of this study are available under <https://osf.io/mjshu/>

SUPPLEMENTARY MATERIAL



Supplementary Figure 1. Pilot data of N=11 participants suggested that conditioned suppression is strongest during the first trial of the PIT phase. That is, participants made significantly less correct responses during the presentation of the CS+ than the new stimulus on trial 1 ($Z=-2.35$, $p=.02$, $d_z=.78$), but not the remaining 5 trials (all p 's > .05). In order to be able to assess the speed of extinction of potential PIT effects, all six trials were retained for the main experiment, but the effect of conditioned suppression only tested on the first trial. During the pilot, the instrumental stimulus was also presented in the PIT phase. This stimulus was not included in the main experiment anymore as it did not serve to test any specific hypotheses.



Supplementary Figure 2. Time between succeeding responses averaged over time-bins of 1000 ms. There was no significant effect of the presentation of the CS+ compared to the new control stimulus on the time between succeeding responses in the beginning of the first CS+ and new stimulus trial (0-1000 ms; $t_{30} = .76$, $p = .46$). As not every participant pressed the buttons more than once within the first 1000 ms of the CS+ (N=38) or the new stimulus presentation (N=40) only N=31 participants contributed to these data.