RUNNING HEAD: PERSISTENT AVOIDANCE

Negative Reinforcement Rate and Persistent Avoidance Following Response-Prevention Extinction

Daniel V. Zuj[1], Weike Xia[1], Keith Lloyd[2], Bram Vervliet[3], and Simon Dymond[1,4,*]

[1] Experimental Psychopathology Lab, Department of Psychology, Swansea University, Singleton Campus, Swansea SA2 8PP, United Kingdom.

[2] Swansea University Medical School, Swansea University, Singleton Campus, Swansea SA2 8PP, United Kingdom.

[3] Laboratory of Biological Psychology, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, 3000 Leuven, Belgium.

[4] Department of Psychology, Reykjavík University, Menntavegur 1, Nauthólsvík, 101 Reykjavík, Iceland.

* Corresponding author.

Abstract

Persistent avoidance may be influenced by prior negative reinforcement rate (i.e., how effective the response is at controlling threat). In clinical settings, the effectiveness of extinction-based methods for treating anxiety-related avoidance may be impacted by prior reinforcement rate. Here, we conducted a laboratory-based treatment study to investigate the persistence of avoidance following response-prevention extinction (RPE) when prior rates of avoidance had been differentially effective at cancelling shock. Participants in three negative reinforcement rate groups (100%, 50%, and 0%) completed a validated avoidance conditioning paradigm involving Pavlovian fear extinction, RPE, and re-extinction phases. It was hypothesised that partially reinforced avoidance rates would lead to diminished resistance to fear extinction following response prevention, compared to continuously- or never-reinforced avoidance. Persistent avoidance was related to prior negative reinforcement rate, with higher rates more resistant to extinction. These findings illustrate the role of reinforcement rate in the persistence of avoidance and may aid understanding of treatment relapse.

Keywords: avoidance; negative reinforcement rate; extinction; response-prevention; anxiety

1. Introduction

Persistent avoidance of real or perceived threat is a central characteristic of anxiety-related disorders (American Psychiatric Association, 2013). Overcoming persistent avoidance represents a considerable challenge to the effectiveness of Pavlovian extinction-based treatments like exposure therapy (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; Lebois, Seligowski, Wolff, Hill, & Ressler, 2019) and the underlying mechanisms are the subject of ongoing debate (Dymond, 2019; Krypotos, Effting, Kindt, & Beckers, 2015; LeDoux, Moscarello, Sears, & Campese, 2017; Lovibond, 2006; San Martín, Jacobs, & Vervliet, 2019). One potential learning mechanism is the prior negative reinforcement rate or effectiveness of avoidance at controlling the aversive event (Xia, Dymond, Lloyd, & Vervliet, 2017). That is, a highly effective avoidance response (i.e., 100% reliable reinforcement rate of shock cancellation) may be more (or less) difficult to treat than a partially effective avoidance response (i.e., 50% reliable reinforcement rate). In clinical settings, the absence of perceived control over threat may contribute to increased vulnerability to anxiety (Gallagher, Bentley, & Barlow, 2014). The role of controllability in avoidant-based coping may therefore be determined, at least in part, by the reinforcement rate of avoidance at minimising potential threat. For example, in social anxiety disorder, turning off one's phone effectively avoids all invitations to social events (i.e., a high reinforcement rate), while on the other hand, not making eye contact with other guests at a social gathering to avoid bein talked to is not always effective (i.e., a lower reinforcement rate). The present study sought to investigate the role of negative reinforcement rate in the persistence of avoidance with the aim of guiding exposure treatment development and mitigating treatment relapse.

Variants of the Pavlovian threat (fear) conditioning paradigms are widely used to study the treatment-related mechanisms of avoidance (Dymond, 2019; LeDoux et al., 2017;

Zuj & Norrholm, 2019). Such paradigms first involve the pairing of a neutral stimulus with an aversive unconditional stimulus (US; e.g., electric shock) such that the cue, now termed a conditional stimulus (CS+), reliably predicts threat. A second neutral stimulus is followed by the absence of the US, thereby becoming a learned safety cue (CS-). Then, to model how avoidance is acquired and maintained involves adding the opportunity to perform a discrete response which, if made in the presence of CS+, postponses upcoming US delivery, and which is deemed unnecessary if made in the presence of a CS-. Withholding US deliveries on fear extinction trials permits investigation of how avoidance may persist but which should, ultimately, extinguish.

The impact of prior reinforcement rate on the persistence of avoidance is germane to the widely studied partial reinforcement extinction effect (PREE), whereby partially reinforced behaviour generally shows greater resistance to extinction than continuously reinforced behaviour (Harris, Kwok, & Gottlieb, 2019; Williams, 2019). Partial reinforcement of avoidance increases resistance to extinction in nonhumans (Galvani, 1971; Marsh & Paulson, 1968), when implemented in concert with other learning related task parameters (e.g., delayed CS termination). In humans, we recently found that higher reinforcement rates during acquisition produced more sustained avoidance responding (keeping CS termination and other factors constant) during an avoidance test phase where the US was withheld, while lower reinforcement rates led to a decrease in responding as extinction progressed (Xia et al., 2017). Groups of participants varied in the effectiveness of avoidance at preventing shock, with rates ranging from 100% reliability (i.e., all shocks could be avoided, or continuous reinforcement rate) to 75%, 50%, 25% and 0% (i.e., no shocks could be avoided, or a range of partial reinforcement rates). Threat expectancy was inversely related to reinforcement rate with the 0%, 25% and 50% groups showing a decline in expectancy ratings, while the 75% and 100% groups' ratings of the likelihood of the US

exhibited a more moderate decline during extinction. No discernible effects of prior reinforcement rate on SCR were detected. Xia et al. (2017) highlighted that Pavlovian fear-based extinction of avoidance may be driven, at least in part, by prior controllability of the aversive event, and also found minimal evidence for the PREE.

Traditional theoretical accounts of these contrasting PREE findings in humans and nonhumans refer to the frustration produced by non-reinforced trials (Amsel, 1967) or the sequential memory basis of non-reinforcement effects on subsequent reinforced trials (Capaldi, 1966). Recent accounts have emphasised the role of reinforcement rate, which when equated between partially and continuously reinforced responses, does still produce the PREE (Bouton, Woods, & Todd, 2014; Chan & Harris, 2017; see also, Seitz, Stolyarova, & Blaisdell, 2019).To date, however, the majority of research has tended to investigate the PREE in the context of Pavlovian conditioning or operant positive reinforcement, with minimal attention given to negative reinforcement rate effects. To better understand how prior negative reinforcement rate may be related to the subsequent persistence of avoidance requires designs that incorporate additional extinction testing procedures. The objectives, therefore, of the present study were to contribute to the literature on the PREE in negatively reinforced avoidance behaviour and examine the role of negative reinforcement rate in the persistence of avoidance before and after a period of response prevention with fear extinction (RPE; Rodriguez-Romaguera et al., 2016).

In laboratory-based treatment studies, RPE involves making the avoidance response unavailable (usually by removing the onscreen cue for avoidance availability): avoidance behaviour is predicted to decrease as participants discover that the aversive event no long occurs (Baum, 1966; Dymond, 2019; Rodriguez-Romaguera et al., 2016; van den Hout, Engelhard, Toffolo, & van Uijen, 2011; Vervliet & Indekeu, 2015; Vervliet, Lange, & Milad, 2017). This RPE procedure resembles a component of exposure therapy for persistent

avoidance, where a central aim is to prevent or reduce the availability of opportunities to engage in avoidance (Abramowitz, Deacon, & Whiteside, 2019; Blakey & Abramowitz, 2016; Dymond, 2019; Rodriguez-Romaguera, Greenberg, Rasmussen, & Quirk, 2016). Response prevention with extinction (RPE) thus provides confirmatory opportunities that avoidance is unnecessary since shock is withheld, which is an integral part of exposure with response prevention treatment for anxiety. Vervliet and Indekeu (2015) found that threat expectancy rapidly recovered when avoidance was available again after RPE. Similarly, Vervliet et al. (2017) noted that avoidance decreased following RPE but did not subsequently recover, while threat expectancy, skin conductance response (SCR) and relief/pleasantess ratings all declined during re-extinction testing following an initial increase when avoidance had been possible. Little is known about the role of prior negative reinforcement rate in persistent avoidance studied using these three-phase extinction designs (extinction, RPE, and re-extinction). The present study therefore investigated persistence of partially reinforced avoidance using a clinically analogous RPE procedure in which the availability and non-availability of the avoidance response was manipulated.

Participants in three reinforcement rate groups (100%, 50%, and 0%) completed a validated avoidance learning paradigm involving extinction, RPE, and re-extinction. It was hypothesised that partially reinforced avoidance (i.e., the 50% reinforcement group) would show more resistance to extinction following response prevention, compared to the full or zero reinforcement of avoidance (i.e., the 100% and 0% reinforcement rate groups, respectively). We expected minimal avoidance responding in the 0% group from the outset, with a decrease across extinction test phases and a sustained level of responding in the 100% group across all phases. This effect will be expressed as a reduction in avoidance during re-extinction, compared to pre-response prevention levels. Further, exploratory analyses will examine any potential individual differences in persistence of avoidance.

2. Methods

2.1. Participants

One hundred and three participants (68 females, 35 males, M = 20.7 years, SD = 2 years, range 18-28 years) were recruited from Swansea University. Sample size was based on our previous research (Xia et al., 2017), and a post-hoc sensitivity analysis using *G*Power 3.1* (Faul, Erdfelder, Lang, & Buchner, 2007) with Power (1 – β) = 0.80, 3 groups and 2 measurements, computed the smallest effect size we should find of Cohen's f = 0.16. Exclusion criteria consisted of: (a) age range outside of 18-40 years old, (b) history of any physical condition possibly affected by the electrocutaneous stimulus (e.g., epilepsy, heart-related conditions and severe migraines), and (c) current use of psychoactive medication. Four participants (3.9%) were excluded as they failed to comply with task instructions, which resulted in 99 participants' data eligible for further analysis. Participants were randomly assigned to one of the three groups: the *100% reinforcement rate* group (n = 33; 23 females, 10 males), *50% reinforcement rate* group (n = 33; 19 females, 14 males), and the *0% reinforcement rate* group (n = 33; 24 females, 9 males). There were no significant between-group differences in mean age and questionnaire scores (all F's < 1.86, all p's > .05). Written consent was obtained at the outset and participants were compensated with either course credits or a £10 shopping voucher. This study protocol was approved by the Department of Psychology Research Ethics Committee, Swansea University.

2.2. Apparatus and Stimuli

Two visual stimuli (grayscale coloured square and triangle, respectively, presented against a white background in the centre of the screen, 2 × 2.5 cm) were used as CS+ and CS-, respectively (counterbalanced across participants) and text consisting of "the spacebar is now available" was used as the onscreen avoidance cue (Figure 1). Stimuli were presented on a 17″ computer screen with a 60 Hz refresh rate in a task programmed in *OpenSesame*

(Mathôt, Schreij, & Theeuwes, 2012).

Insert Figure 1 About Here

The US was a 250 ms duration electric shock generated using a STM200 stimulator (BIOPAC Systems, Santa Barbara, USA) administered through a surface electrode (MLADDF30 bar electrode with two, 9 mm contacts spaced 30 mm apart). Electrode gel was applied to the right forearm and the electrode held in place with a Velcro band. Shock was individually calibrated at the beginning of the session. The current was initially set at 35 mV and increased or decreased in steps of 2.5 mV (the maximum was 100 mV). Participants were asked to report the intensity of the shock in terms of how uncomfortable they found it. When a shock level was deemed "uncomfortable but not painful" twice consecutively, it was used for that participant. Skin conductance was measured through two Ag/AgCl electrodes coated with non-hydrating gel attached to the middle phalanges on the index and middle fingers and interfaced with the MP150 (BIOPAC Systems, Santa Barbara, USA). The SCR signal was sampled at 1000Hz with a notch filter of 10Hz.

For all trials, a fixation cross was presented for 2 s followed by the CS (15 s). Trials commenced with a 5 s presentation of the CS only, to allow for SCR recording. Following this, the avoidance cue was presented onscreen below the CS for 2 s, during which time participants could make an avoidance response (i.e., spacebar press). Next, a Visual Analogue Scale (VAS) ranging from 0 ("*I certainly expect no shock*") to 100 ("*I certainly expect a shock*") was shown underneath the CS as a self-report measure of threat expectancy (Figure 1). Participants responded by sliding the mouse along the VAS and pressing the left mouse button to confirm their choice. The threat expectancy scale was removed following a rating or on CS termination, whichever happened first (see Figure 1). The intertrial interval (ITI) ranged between 6 and 9 s. CSs were presented pseudorandomly with no more than two consecutive presentations of the same trial type. In all phases after *threat conditioning*, trials

were blocked in groups of 8 (4 CS+ and 4 CS-). After each block, participants rated how fearful they found the CSs to be using a scale that ranged from 0 ("*Not at all fearful")* to 100 (*"Very fearful"*).

Insert Figure 1 about here

2.3. Procedure

Participants first completed the consent form and questionnaires (Supplementary Materials) before shock calibration and attachment of the skin conductance electrodes. The experiment consisted of five phases in a single, one-day session: *threat conditioning, avoidance conditioning, extinction, response prevention and extinction,* and *re-extinction* (see Supplementary Materials for a copy of the instructions used).

During *threat conditioning*, each CS was presented twice in a quasi-random order; all CS+ trials were coupled with the US, which occurred at stimulus offset. The US never followed CS- presentations.

The *avoidance conditioning* phase consisted of 8 CS+ and 8 CS- trials. In this phase, the avoidance cue appeared 5 s after CS onset. In the 100% group, participants could cancel upcoming shock after every CS+ trial if they made the avoidance response during each trial. For the 50% group, avoidance was partially effective at cancelling upcoming shock (i.e., there were 4/8 avoidable trials) and there were zero avoidable CS+ trials in the 0% group. The absence of avoidance responding on CS+ trials was always followed by shock, regardless of group.

During the *extinction* phase, CS+ and CS- were each now presented 20 times and shock was withheld on all trials (avoidance and non-avoidance) for all groups. In the *response-prevention and extinction* (RPE) phase, participants in all reinforcement groups were no longer shown the avoidance cue making the avoidance response unavailable. CS+ and CS- were presented 8 times each and the US was not presented. Finally, the *re-extinction*

phase reintroduced the avoidance cue (and hence the availability of avoidance), and similar to before, no shock was given regardless of responding. CS+ and CS- were presented 4 times each.

After the final trial, SCR and shock electrodes were removed, and participants were debriefed and compensated. The task took approximately 45 min, and the total session lasted approximately 60 min.

2.4. Data analysis

Skin conductance responses (SCRs) were calculated using *AcqKnowledge* software (BIOPAC Systems, Santa Barbara, USA) as the maximum response to occur within 0.5–5 s post-CS onset and were range-corrected for each participant and square-root transformed prior to statistical analyses. For the total sample (N = 103) there were six participants with zero SCR responses (5.8%). After excluding four participants for failing to follow the task instructions, 14 of the remaining 99 participants (14.1%) were defined as SCR non-responders (> 90% zero responses; Marin et al., 2019; Xia et al., 2017) and excluded from analyses of SCR, leaving a total sample size for these analyses of N = 85 (100% group, n = 30; 50% group, n = 28; and 0% group, n = 27). Avoidance proportion data was scored as the percentage of trials on which avoidance occurred per phase and per stimulus. A mixed ANOVA was used to analyse fear ratings by CS, block and phase. To assess the immediate effect of the removal of the avoidance cue during *response-prevention and extinction*, single trial comparisons between the last trial of *extinction* and the first trial of *response-prevention and extinction* were made for threat expectancy ratings and SCR. Similarly, to assess the reintroduction of the avoidance cue, the last trial of the *response-prevention and extinction* phase and the first trial of *re-extinction* were compared, as recommended for associative learning studies with between-phase reinstatement cues (Haaker, Golkar, Hermans, & Lonsdorf, 2014). Finally, proportion of avoidance responses were compared between the

extinction phase and the *re-extinction* phase. Greenhouse-Geisser corrections were applied where sphericity was not met, effect sizes are reported as partial eta-squared (η_p^2), and the alpha level was set to $\alpha = .05$ unless otherwise stated.

Insert Figures 2, 3, 4 and 5 About Here

3. Results

3.1. Threat conditioning

Ratings. 3 (Group) × 2 (CS) × 2 (Trial) mixed ANOVA was conducted for threat expectancy ratings, which confirmed differential threat conditioning (see Figure 3). For threat expectancy ratings, the Group × CS × Trial interaction was not significant ($F_{(2,53)} = 0.39$, $p = .680$, $\eta_p^2 = .014$) and neither was the CS × Trial interaction ($F_{(1,53)} = 2.98$, $p = .090$, $\eta_p^2 = .053$). This is not surprising as participants were instructed on the CS-US contingency at the beginning of this phase. There was, importantly, a significant main effect of CS ($F_{(1,53)} = 179.55$, $p < .001$, $\eta_p^2 = .772$) with higher expectancy ratings to the CS+ relative to the CS-, indicating that participants read and understood the instructions, and displayed differential threat expectancy (see Figure 3). There were no significant main effects of interactions involving group (all p's > .05).

After *threat conditioning,* there was a significant main effect of CS ($F_{(1,96)} = 486.88$, $p < .001$, $\eta_p^2 = .84$), with participants reporting greater fear ratings to the CS+ ($M = 70.03$, $SD = 26.09$) than the CS- ($M = 7.94$, $SD = 12.11$). Although there was no significant main effect of Group ($F_{(2,96)} = 0.54$, $p = .584$, $\eta_p^2 = .01$), there was a significant Group × CS interaction ($F_{(2,96)} = 3.41$, $p = .034$, $\eta_p^2 = .07$). Post-hoc Bonferroni-corrected tests of simple main effects revealed that, although all groups displayed significantly higher fear ratings towards the CS+ than the CS- (all p's < .001), this difference was largest for the 100% reinforcement rate group ($M_{DIFF} = 69.23$, $SEM = 4.87$), with smaller differences for the 50% ($M_{DIFF} = 65.20$, $SEM = 4.87$) and 0% groups ($M_{DIFF} = 51.82$, $SEM = 4.87$).

Skin conductance. Similarly for SCR amplitude, there was no significant Group × CS × Trial interaction ($F_{(1,82)} = 0.22$, $p = .805$, $\eta_p^2 = .005$) or CS × Trial interaction ($F_{(1,82)} = 0.67$, $p = .415$, $\eta_p^2 = .008$). There was, however, a significant main effect of CS ($F_{(1,82)} = 42.17$, $p < .001$, $\eta_p^2 = .340$) with greater SCR amplitude to the CS+ relative to the CS- (see Figure 5). There were no further main effects or interactions involving group (all p's > .05).

3.2. Avoidance conditioning

Proportion of avoidance. A 3 (Group) × 2 (CS) mixed ANOVA revealed a significant main effect of CS ($F_{(1, 96)} = 84.25$, $p < .001$, $\eta_p^2 = .47$), with greater proportion of avoidance to the CS+ ($M = 71.72$, $SD = 36.72$) than the CS- ($M = 35.73$, $SD = 43.20$). The Group × CS interaction was not significant ($F_{(2, 96)} = 2.86$, $p = .062$, $\eta_p^2 = .06$) and nor was the main effect of Group ($F_{(2, 96)} = 2.86$, $p = .062$, $\eta_p^2 = .06$).

Ratings. Figures 3 and 4 show differences in threat expectancy and fear ratings, respectively, for the CS+ throughout the *avoidance conditioning* phase. Both threat expectancy and fear ratings revealed a significant three-way interaction (threat expectancy: $F_{\text{Group x CS x Trial}(8.97, 349.77)} = 7.08$, $p < .001$, $\eta_p^2 = .154$; fear: $F_{\text{Group x CS x Block}(2, 96)} = 8.62$, $p < .001$, $\eta_p^2 = .15$). Analysing threat expectancy per CS revealed a significant Group × Trial interaction for the CS+ ($F_{(9.58, 402.48)} = 8.07$, $p < .001$, $\eta_p^2 = .16$). Planned contrasts revealed that all groups differed (p's < .05) with ratings highest in the 0% group and lowest in the 100% group. Bonferroni-corrected pairwise comparisons revealed that, from Trial 2 onwards, ratings for the 0% group were significantly higher than the 100% group (p's < .001), but were not significantly different than the 50% group (p's < .08). For the 50% group, threat expectancy was significantly higher than for the 100% group from trial 4 onwards. Figure 3 illustrates the steeper increase in expectancy in the 0% group compared to the 50% group. Planned contrasts of fear ratings for the CS+ showed no group differences in block 1 (p's >

.05). In block 2, both the 0% and 50% groups reported higher fear ratings than the 100% group (p's < .001) but were not different from each other (p > .05).

Figure 4 shows that fear ratings decreased in the 100% group over time but remained stable in both the 0% and 50% group. For CS-, there was only a main effect of Trial ($F_{(4.62, 410.94)} = 4.34$, $p < .01$, $\eta_p^2 = .05$) indicating that threat expectancy ratings for the CS- were unaffected by reinforcement rate. Similarly, planned contrasts revealed that fear ratings did not differ between groups for the CS- in either block 1 or block 2 (p's > .05).

Skin conductance. The ANOVA revealed a significant Group × CS × Trial interaction ($F_{(12.14, 497.59)} = 1.89$, $p = .03$, $\eta_p^2 = .05$), and follow-up analyses showed that for the CS+ trials, there were no significant simple main effects of Group ($F_{(2, 82)} = 0.71$, $p > .05$), Trial ($F_{(5.86, 480.64)} = 1.17$, $p > .05$) or a Group × Trial simple interaction effect ($F_{(11.72, 480.64)} = 0.96$, $p > .05$). On CS- trials, however, there was a significant Group × Trial simple interaction ($F_{(11.80, 483.9)} = 2.31$, $p < .01$, $\eta_p^2 = .05$), and trend analyses showed a significant quadratic trend for this interaction ($p < .05$). Figure 5 illustrates this quadratic trend with SCR to CS- trials initially increasing for the 0% group and subsequently declining, while the 50% and 100% groups showed a rapid increase during the final two trials of the phase. Planned contrasts showed no significant group differences in SCR to both CS+ and CS- (p's > .05).

3.3. Extinction

Proportion of avoidance. During the *extinction* phase, a 3 (Group) × 2 (CS) mixed ANOVA revealed a significant main effect of CS ($F_{(1, 96)} = 57.06$, $p < .001$, $\eta_p^2 = .37$), with significantly greater proportion of avoidance to the CS+ ($M = 58.13$, $SD = 43.93$) than the CS- ($M = 28.23$, $SD = 41.84$). This main effect was superseded by a significant Group × CS interaction ($F_{(2, 96)} = 3.90$, $p = .024$, $\eta_p^2 = .08$). Bonferroni-corrected simple main effects found that although each group displayed significantly greater avoidance behaviour to the CS+ relative to the CS-, this difference was greatest for the 100% reinforcement group

(M_{DIFF} = 39.39, *SEM* = 6.86, $p < .001$) compared to the 50% group (M_{DIFF} = 35.91, *SEM* = 6.86, $p < .001$), with the smallest difference found for the 0% reinforcement rate group (M_{DIFF} = 14.39, *SEM* = 6.86, $p = .038$). This effect can be seen in Figure 2. Finally, there was no significant main effect of Group ($F_{(2, 96)} = 0.45$, $p = .642$, $\eta_p^2 = .01$).

Ratings. Figures 3 and 4 show a clear differential decline in ratings between groups over time, and this was confirmed by significant Group × CS × Trial interactions for threat expectancy ratings ($F_{(12.27, 506.98)} = 4.79$, $p < .001$, $\eta_p^2 = .12$) and fear ratings ($F_{(4.28, 205.19)} = 2.87$, $p = .02$, $\eta_p^2 = .06$). As such, analyzing threat expectancy per CS revealed a significant Group × Trial interaction ($F_{(12.09, 507.85)} = 8.76$, $p < .001$, $\eta_p^2 = .17$) for the CS+ only. Figure 3 shows much steeper declines in threat expectancy in the 0% and 50% groups compared to the 100% group; this difference most likely drove the interaction effect. To assess whether there was also a difference between the rate of decline between the 0% and 50% group, both groups were directly compared in a 2 (Group) × 20 (Trial) repeated measures ANOVA. There was a significant Group × Trial interaction ($F_{(6.71, 368.9)} = 5.52$, $p < .001$, $\eta_p^2 = .09$) and further trend analyses showed a significant linear trend ($p < .05$) indicating that the decrease in threat expectancy was faster in the 0% group compared to the 50% group. Planned contrasts revealed that throughout the *extinction* phase, threat expectancy ratings for the CS+ were lower in the 100% group compared to both the 0% and 50% groups (p's < .05), while there was no overall difference between 0% and 50% groups.

ANOVA further revealed a significant Group × Trial interaction ($F_{(4.20, 209.9)} = 2.80$, $p = .03$, $\eta_p^2 = .05$) and Figure 4 shows that during *extinction* there was a general decline in fear ratings. Specifically, the 0% group showed a reduction in ratings to similar levels of the 100% group, while fear ratings in the 50% group remained higher. Post-hoc Bonferroni-corrected tests of simple main effects confirmed this pattern in block 1, while fear ratings between the 50% and 0% groups did not differ ($p > .05$), but both groups' ratings were higher

than the 100% group (p's < .05). However, by block 5, fear ratings in the 50% group were significantly higher than both the 0% and 100% groups (p's < .05), and there was no difference between the 0% and 100% group (p > .05). Together, these results indicate that participants in the 50% group were slower to learn that shock was withheld compared to the 0% group, and remained overall more fearful of the CS+. In comparison, analyses of ratings made in response to the CS- showed a general decrease over trials for threat expectancy ratings ($F_{(5.12, 404.4)} = 3.37$, $p < .01$, $\eta_p^2 = .04$) and fear ratings ($F_{(1.70, 169.69)} = 4.73$, $p = .01$, $\eta_p^2 = .05$), with no other significant effects (F's ≤ 1.49, p's > .05).

Skin conductance. Figure 5 shows a general decline in SCR over time, as shown in the 3 (Group) × 2 (CS) × 20 (Trial) mixed ANOVA which found a significant main effect of Trial ($F_{(10.43, 854.9)} = 30.51$, $p < .001$, $\eta_p^2 = .23$). There was a significant Group × CS interaction ($F_{(2, 82)} = 3.43$, $p = .04$, $\eta_p^2 = .08$). Post-hoc tests revealed that for the CS+ there were no group differences (p's > .05); however, SCR to the CS- was significantly higher in the 0% group than in the 100% group ($p < .01$). There were no further significant main effects or interactions involving group (p's > .05).

3.4. Response prevention and extinction

Ratings. Threat expectancy for the CS+ immediately increased for each group at the beginning of the response prevention and extinction phase, but not for the CS- (see Figure 3). This was indicated by the results of the 3 (Group) × 2 (CS) × 2 (Phase) mixed ANOVA, which showed a significant CS × Phase interaction ($F_{(2.71, 232.8)} = 18.60$, $p < .001$, $\eta_p^2 = .18$). Bonferroni-corrected pairwise comparisons further confirmed that threat expectancy ratings increased significantly from the *extinction* phase to the *response prevention and extinction* phase to only the CS+ in all groups (p's < .05), but not the CS- (p's > .05). Fear ratings similarly increased from the onset of the RPE phase ($F_{\text{CS x Phase } (1, 96)} = 7.48$, $p < .01$, $\eta_p^2 = .07$) to the CS+ only (p's < .05), with no significant effects for the CS- (p's > .05).

During RPE, a significant main effect of Trial showed threat expectancy decreased throughout this phase ($F_{(2.74, 233.4)} = 27.85$, $p < .001$, $\eta_p^2 = .25$), but there was no significant Group × CS interaction ($F_{(2, 86)} = 2.77$, $p = .07$, $\eta_p^2 = .06$). Fear ratings showed a similar decrease across groups across trials ($F_{(1, 96)} = 13.48$, $p < .001$, $\eta_p^2 = .12$) and a significant Group × CS interaction effect ($F_{(2, 96)} = 5.34$, $p < .01$, $\eta_p^2 = .10$). Pairwise comparisons showed that fear ratings to the CS+ were higher in the 50% group than the 100% group ($p < .05$), but there were no differences in CS- ratings (p's > .05). Together these results indicate that removal of the avoidance cue led to an overall decrease in threat expectancy, but subjective fear ratings to the CS+ remained higher in the 50% group compared to the other groups.

Skin conductance. Comparing the last trial of *extinction* to the first trial of *response prevention and extinction* revealed no significant main effects (F's ≤ 2.27, p's > .05) or interaction effects (F's ≤ 2.41, p's > .05). Furthermore, there was a significant main effect of Trial ($F_{(2.49, 204.03)} = 5.45$, $p < .05$, $\eta_p^2 = .04$), but no significant main effect of CS ($F_{(1, 82)} = 3.76$, $p = .06$, $\eta_p^2 = .04$) or main effects and interactions involving group (F's < 1, p's > .05). These results indicate a general distinction between CS+ and CS- and decreases in SCR over time regardless of reinforcement rate.

3.5. Re-extinction

Proportion of avoidance. To assess whether the reintroduction of the avoidance cue following *response prevention and extinction* led to persistence of avoidance, a 3 (Group) × 2 (CS) × 2 (Phase) mixed ANOVA compared the proportion of avoidance during *extinction* to those during *re-extinction*. The ANOVA revealed a significant main effect of CS ($F_{(1, 96)} = 44.94$, $p < .001$, $\eta_p^2 = .32$), with greater proportion of avoidance to the CS+ ($M = 51.92$, $SD = 43.52$) than the CS- ($M = 26.36$, $SD = 41.46$). Further, there was a significant main effect of Phase ($F_{(1, 96)} = 20.1$, $p < .001$, $\eta_p^2 = .17$), with a greater proportion of avoidance during the

extinction phase (M = 43.12, SD = 38.03) compared to the *re-extinction* phase (M = 35.10, SD = 40.11). These main effects were superseded by a significant CS × Phase interaction ($F_{(1,96)}$ = 11.4, p = .001, η_p^2 = .11). Bonferroni-corrected simple main effects revealed that the difference in avoidance proportions between *extinction* and *re-extinction* was greater for the CS+ (M_{DIFF} = 12.42, SE = 2.71, p < .001) than the CS- (M_{DIFF} = 3.74, SE = 1.57, p = .019), with general avoidance rates being lower in the *re-extinction* phase compared to *extinction*. Contrary to the primary hypothesis, this effect did not change as a function of reinforcement rate group, as indicated by the non-significant Group × CS × Phase interaction, ($F_{(2, 96)}$ = 1.6, p = .207, η_p^2 = .03) and, importantly, the achieved effect size for this interaction is similar to the lowest effect size computed by the sensitivity analysis. Further, there was a non-significant main effect of Group ($F_{(2, 96)}$ = 0.27, p = .767, η_p^2 = .006), with non-significant Group × CS ($F_{(2, 96)}$ = 2.64, p = .077, η_p^2 = .05) and Group × Phase interactions ($F_{(2, 96)}$ = 0.75, p = .476, η_p^2 = .02).

Ratings. While Figure 3 shows an increase in CS+ threat expectancy ratings in only the 0% group following the reintroduction of the avoidance cue, the 3 (Group) × 2 (CS) × 2 (Phase) mixed ANOVA revealed no significant three-way interaction ($F_{(2, 95)}$ = 2.84, p = .06, η_p^2 = .06). There were no significant main effects or interactions involving group (p's > .05).

Alternatively, fear ratings showed a significant Group × CS interaction ($F_{(2, 96)}$ = 3.61, p < .05, η_p^2 = .07), with Bonferroni-corrected pairwise comparisons revealing that fear of the CS+ was significantly higher than the CS- in the 0% and 50% groups (p's < .05), but not in the 100% group (p > .05).

Skin conductance. There were no significant differences in SCR during the *re-extinction* phase (F's ≤ 1.00, p's > .05).

**** Insert Table 1 About Here ****

3.6. Post-hoc analyses: Avoided vs. Non-avoided trials.

Building on our previous work (Xia et al., 2017), and to determine whether threat expectancy reflected perceived effectiveness of avoidance, threat expectancy ratings on trials with a successful avoidance response (*Avoided*) and no avoidance response (*Non-Avoided*) were compared (Table 1). Due to the unequal sample sizes, one-way ANOVA with weighted means were used comparing Avoided and Non-Avoided trials per phase (further analyses may be found in Supplementary Materials).

Avoidance conditioning. Threat expectancy ratings for Avoided CS+ trials were significantly different between groups ($F_{(2, 81)} = 70.35$, $p < .001$, $\eta_p^2 = .50$) showing a linear increase in ratings from 100% to 0% reinforcement rate (Table 1). Post-hoc tests showed that each group was significantly different from one another, with lower expectancy ratings in higher reinforcement rate groups (p's < .001). There were no group differences in threat expectancy ratings for Non-Avoided CS+ trials, and all CS- trials (all p's > .05).

Extinction. Similarly, threat expectancy ratings differed in the Avoided CS+ trials ($F_{(2, 72)} = 67.14$, $p < .001$, $\eta_p^2 = .48$), but not in the Non-Avoided CS+, Avoided CS- and Non-Avoided CS- trials (all p's > .05). The 100% group made significantly lower threat expectancy ratings than the 50% and 0% groups (p's < .001), with no difference between the 50% and 0% groups ($p > .05$).

Re-extinction. Threat expectancy ratings also differed on Avoided CS+ trials when the opportunity to avoid was made available again ($F_{(2, 50)} = 4.33$, $p < .05$, $\eta_p^2 = .05$); however, this was no longer significant in follow-up comparisons (p's > .05). There were no further differences on all remaining trial types (all p's > .05).

3.6.1 Supplementary analysis: Fear ratings. Supplementary analysis was conducted to examine fear ratings made of the CS+ and CS- at the end of *threat conditioning* between the participants who avoided every CS+ trial during *avoidance conditioning* ($n = 47$) and the participants who did not avoid any CS+ trial ($n = 15$). It was found that the avoiders reported

significantly higher fear ratings at the end of threat conditioning (M = 76.26, SD = 23.24) than the non-avoiders (M = 40.75, SD = 30.57), $t(19.4) = -4.13$, $p = .001$, equal variances not assumed. There was no significant difference between avoiders and non-avoiders in fear ratings of the CS- at the end of threat conditioning, $t(16.4) = 1.02$, $p = .321$, equal variances not assumed.

In summary, threat expectancy during *avoidance conditioning* was higher in lower reinforcement rate groups. The 50% and 0% reinforcement rate groups maintained higher threat expectancy than the 100% group during *extinction* when the US was withheld. Following RPE, these between-group differences were not evident during the *re-extinction* phase. Finally, prior differences in fear ratings made by those participants who were subsequently categorised as avoiders or non-avoiders may have influenced avoidance conditioning, with higher self-reported fear driving increased avoidance.

4. Discussion

In the current study, we manipulated the negative reinforcement schedule of avoidance in a validated conditioning and extinction task with an intervening response prevention phase. It was hypothesised that partially reinforced avoidance (i.e., the 50% reinforcement group) would be associated with a reduction in responding following the response prevention phase, relative to full (100%) and zero negative reinforcement rates. While avoidance was significantly reduced during re-extinction compared to the extinction phase, providing partial support for the hypothesis, it did not differ as a function of reinforcement rate. Specifically, the 100% group continued to show higher avoidance rates than the 50% group who, in turn, showed higher avoidance rates than the 0% group, although between-group comparisons were non-significant. The rate of avoidance following the

response prevention phase did, however, differ as a function of CS, with understandably greater avoidance in the presence of the CS+ relative to the CS-.

The present findings replicated a trend that we previously identified of a gradient-like profile in the acquisition of avoidance in the presence of the CS+ (Xia et al., 2017). Prior effectiveness (negative reinforcement rate) determined avoidance responding, with a higher level in the 100% reinforcement group followed by 50% and 0% groups, respectively. Threat expectancy corroborated these behavioural findings, with a steady decline across trials in the likelihood of shock in the 100% group and an increasing trend evident early on in acquisition for the 0% group, while the 50% group's expectancy ratings occupied an intermediate level between the two. Taken together, these findings replicate that reinforcement rate is directly related to both the learning rate of avoidance responding and to ratings of its effectiveness at canceling shock. Similarly, during extinction we found that this gradient-like profile persisted, with a decrease in avoidance compared to learning levels. Threat expectancy during this phase, however, showed a clear differential pattern between the groups: ratings of the CS+ decreased more rapidly in the 0% reinforcement group compared to the 50% group, while both groups displayed a sharper decline compared to particpants in the 100% group (Xia et al., 2017). During response prevention and extinction, there was an immediate and pronounced increase in CS+ threat expectancy followed by an overall decrease as participants learned that shock was withheld. This initial increase and subsequent decrease was of a similar magnitude in all groups, with the 50% group's expectancy level remaining consistently higher than the others. This supports the findings of Vervliet and Indekeu (2015) who also noted a similar trend in ratings made of avoidable and unavoidable CS+s. Together these results indicate that removal of the avoidance cue led to an overall decrease in ratings but expectancy and fear of the CS+ remained relatively higher in the 50% group compared to the other groups. Unfortunately, as response rate was not recorded during the response

prevention and extinction phase (cf. Dymond, 2019), we were unable to determine any possible effects of reinforcement rate on response disruption. Persistent avoidance was evident, albeit at reduced levels compared to the first extinction phase, with higher avoidance rates predicted by higher reinforcement rates but not significantly. Threat expectancy remained low, however, with between-group trends persisting across both extinction phases.

The present study may have important clinical implications for the treatment of anxiety-related avoidance with response prevention techniques (Abramowitz et al., 2019; Lebois et al., 2019). Specifically, we found that higher rates of avoidance reinforcement during acquisition result in a level of resistance to extinction and return of avoidance following response prevention, compared to partial or zero avoidance reinforcement. In a clinical setting, these findings indicate that the learning history of avoidance acquisition, and its previous success at cancelling negative outcomes, may be an indicator of the likelihood of treatment effectiveness. That is, we suggest that individuals with a previous learning history with fully effective avoidance behaviours may be more resistant to treatment change. This means that in the therapeutic setting, it may be advantageous for individuals with problematic avoidance to experience partially-reinforced learning exemplars whereby the aversive event may still occur prior to subsequent extinction-based treatment. This would allow for disconfirming opportunities whereby the aversive event is occasionally withheld, hence undermining the ubiquity of perceived avoidance reliability and enhancing extinction.

The study has a number of limitations. First, the sequential design employed meant that participants in the three reinforcement rate groups were likely to be responding at different levels of avoidance (and non-avoidance) prior to the onset of the common test phases. Any resulting between-group differences during extinction and RPE may have been impacted by the potential imbalance in obtained negative reinforcement rates. Future studies employing alternative designs, such as single case experimental designs with predetermined

acquisition criteria (e.g., Lejeuz, O'Donnell, Wirth, Zvolensky, & Eifert, 1998), may be helpful in further disentangling the effects of prior reinforcement on persistent avoidance. Second, we used an avoidance response (spacebar press) that required a low response effort (Courtney & Perone, 1992; Friman, 1995) and which resulted in no monetary cost to the participant (Krypotos, Vervliet, & Engelhard, 2018). Low effort avoidance and related safety behaviours can be quite persistent in clinical disorders, such as carrying anti-anxiety medications in case of a panic attack, and may thus be difficult to extinguish (Vervliet & Indekeu, 2015). While the clinical relevance of a discrete avoidance response which was easy and low cost to perform remains to be seen, it is important to note that across all groups, the avoidance response was in effect a fixed ratio (FR) 1 schedule of (continuous) reinforcement which then initiated one of the specified rates of shock cancellation (100%, 50% or 0%). Further adaptations of the paradigm could add an increased reinforcement schedule requirement, with and without a monetary cost, in order to improve validity and enhance translational relevance (Krypotos et al., 2018; Pittig, Wong, Glück, & Boschet, 2020; Poling, 2010). Third, we did not measure avoidance responding from the outset of the response-prevention extinction procedure. Doing so in the future will permit analysis of the time-course of avoidance extinction and help identify possible factors influencing the decision to avoid or not to avoid from the first trial onwards. It will also help to unambiguously determine whether avoidance had in fact extinguished to zero or near-zero rates (Dymond, 2019). Fourth, to better understand the role of individual differences and within-session performance on persistent avoidance, future studies should correlate the number of trials on which participants failed to made the avoidance response during an extinction phase with subsequent fear and avoidance behaviour. Doing so would greatly aid interpretation of sequential effects across learning and test phases in designs like this. Fifth, the SCR analysis resulted in the exclusion of 14% of participants as non-responders. Considerable

heterogeneity exists in the threat conditioning literature as to the exclusion criteria applied to SCR and other learning data. Future research would clearly benefit from a consensus view of how best to handle such outcomes (Lonsdorf et al., 2020; Ney et al., 2020). Sixth, fear ratings following each trial block did not specify whether participants should rate the fear they would experience if the avoidance response was available or unavailable. Future research should therefore investigate the role of avoidance availability on the modulation of self-reported fear (Dymond, Shlund, Roche, Whelan, Richards, & Davies, 2011). Finally, to initiate contact with the avoidance response during avoidance conditioning we partially instructed participants that a response was required to cancel upcoming shock rather than merely expose participants to the task without any specific instructions (see Supplementary Materials). Given the emerging focus on the effect of task-related instructions in fear and avoidance conditioning research (Mertens, Boddez, Sevenster, Engelhard, & De Houwer, 2018), it would be worthwhile investigating the extent to which instructions are responsible for the findings of this study.

In conclusion, the present findings demonstrate, for the first time, clear partial reinforcement effects during the acquisition and extinction of avoidance using a *response prevention extinction procedure* and the persistence of avoidance duing *re-extinction*.

References

Abramowitz, J. S., Deacon, B. J., & Whiteside, S. P. H. (2019). *Exposure Therapy for Anxiety: Principles and Practice* (2nd ed.). New York: Guilford.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: Author.

Amsel, A. (1967). Partial reinforcement effects on vigor and persistence: Advances in frustration theory derived from a variety of within-subjects experiments. *Psychology of Learning and Motivation, 1*, 1-65. doi:10.1016/S0079-7421(08)60512-5

Baum, M. (1966). Rapid extinction of an avoidance response following a period of response prevention in the avoidance apparatus. *Psychological Reports, 18*, 59-64. doi:10.2466/pr0.1966.18.1.59

Blakey, S. M., & Abramowitz, J. S. (2016). The effects of safety behaviors during exposure therapy for anxiety: Critical analysis from an inhibitory learning perspective. *Clinical Psychology Review, 49*, 1-15. doi:10.1016/j.cpr.2016.07.002

Bouton, M. E., Woods, A. M., & Todd, T. P. (2014). Separation of time-based and trial-based accounts of the partial reinforcement extinction effect. *Behavioural Processes, 101*, 23–31. doi:10.1016/j.beproc.2013.08.006

Capaldi, E. J. (1966). Partial reinforcement: A hypothesis of sequential effects. *Psychological Review, 73*(5), 459–477. doi:10.1037/h0023684

Chan, C. K. J., & Harris, J . A. (2017). Extinction of Pavlovian conditioning: The influence of trial number and reinforcement history. *Behavioural Processes, 141*, 19-25. doi:10.1016/j.beproc.2017.04.017

Courtney, K., & Perone, M. (1992). Reductions in shock frequency and response effort as factors in reinforcement by timeout from avoidance. *Journal of the Experimental Analysis of Behavior, 58*, 485-496. doi:10.1901/jeab.1992.58-485

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. doi:10.1016/j.brat.2014.04.006

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

Dymond, S., Schlund, M., Roche, B., Whelan, R., Richards, J., & Davies, C. (2011). Inferred threat and safety: Symbolic generalization of human avoidance learning. *Behaviour Research and Therapy*, *49*, 614-621.

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*, 175-191.

Flores, A., López, F. J., Vervliet, B., & Cobos, P. L. (2018). Intolerance of uncertainty as a vulnerability factor for excessive and inflexible avoidance behavior. *Behaviour Research and Therapy, 104*, 34-43. doi:10.1016/j.brat.2018.02.008

Friman, P. C. (1995). Making life easier with effort: Basic findings and applied research on response effort. *Journal of Applied Behavior Analysis, 28*, 583-590. doi:10.1901/jaba.1995.28-583

Gallagher, M. W., Bentley, K. H., & Barlow, D. H. (2014). Perceived control and vulnerability to anxiety disorders: A meta-analytic review. *Cognitive Therapy and Research*, *38*(6), 571-584.

Galvani, P. F. (1971). The effects of partial reinforcement on the acquisition and extinction of avoidance behavior in gerbils. *Psychonomic Science, 24*, 242-244. doi:10.3758/BF03331810.

Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014). A review on human reinstatement studies: an overview and methodological challenges. *Learning and Memory, 21*, 424-440. doi:10.1101/lm.036053.114

Harris, J. A., Kwok, D. W. S., & Gottlieb, D. A. (2019). The partial reinforcement extinction effect depends on learning about nonreinforced trials rather than reinforcement rate. *Journal of Experimental Psychology: Animal Learning and Cognition, 45*(4), 485–501. doi:10.1037/xan0000220

Hunt, C., Cooper, S. E., Hartnell, M. P., & Lissek, S. (2017). Distraction/Suppression and Distress Endurance diminish the extent to which generalized conditioned fear is associated with maladaptive behavioral avoidance. *Behaviour Research and Therapy, 96*, 90-105. doi:10.1016/j.brat.2017.04.013

Krypotos, A.-M., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: A review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience, 9*(189). doi:10.3389/fnbeh.2015.00189

Krypotos, A.-M., Moscarello, J. M., Sears, R. M., LeDoux, J. E., & Galatzer-Levy, I. (2018). A principled method to identify individual differences and behavioral shifts in signaled active avoidance. *Learning and Memory, 25*, 564-568. doi:10.1101/lm.047399.118

Krypotos, A.-M., Vervliet, B., & Engelhard, I. M. (2018). The validity of human avoidance paradigms. *Behaviour Research and Therapy, 111*, 99-105. doi:10.1016/j.brat.2018.10.011

Lebois, L. A. M., Seligowski, A. V., Wolff, J. D., Hill, S. B., & Ressler, K. J. (2019). Augmentation of extinction and inhibitory learning in anxiety and trauma-related disorders. *Annual Review of Clinical Psychology, 15*, 257-284. doi:10.1146/annurev-clinpsy-050718-095634

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*, 24-36. doi:10.1038/mp.2016.166

Lejuez, C. W., O'Donnell, J., Wirth, O., Zvolensky, M. J., & Eifert, G. H. (1998). Avoidance of 20% carbon dioxide-enriched air with humans. *Journal of the Experimental Analysis of Behavior, 70*, 79-86.

Lommen, M. J. J., Engelhard, I. M., & van den Hout, M. A. (2010). Neuroticism and avoidance of ambiguous stimuli: Better safe than sorry? *Personality and Individual Differences, 49*, 1001-1006. doi:https://doi.org/10.1016/j.paid.2010.08.012

Lonsdorf, T. B., et al. (2020). Navigating the garden of forking paths for data exclusions in fear conditioning research. *eLife*, *8*, e52465. doi: 10.7554/eLife.52465

Lovibond, P. F. (2006). Fear and avoidance: An integrated expectancy model. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and Learning: From Basic Processes to Clinical Implications* (pp. 117-132). Washington DC: American Psychological Association.

Marin, M. F., Barbey, F., Rosenbaum, B. L., Hammoud, M. Z., Orr, S. P., & Milad, M. R. (2019). Absence of conditioned responding in humans: A bad measure or individual differences? *Psychophysiology, 57*(1), e13350. doi: 10.1111/psyp.13350

Marsh, G., & Paulson, N. (1968). The effect of partial reinforcement on extinction in avoidance conditioning. *Psychonomic Science, 12*, 39-40. doi:10.3758/bf03331178

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314-324. doi:10.3758/s13428-011-0168-7

Mertens, G., Boddez, Y., Sevenster, D., Engelhard, I. M., & De Houwer, J. (2018). A review on the effects of verbal instructions in human fear conditioning: Empirical findings, theoretical considerations, and future directions. *Biological Psychology, 137*, 49-64. doi:10.1016/j.biopsycho.2018.07.002

Ney, L. J., Laing, P. A .F., Steward, T., Zuj, D. V., Dymond, S., & Felmingham, K. L. (2020). Inconsistent analytic strategies reduce robustness in fear extinction via skin conductance response. *Psychophysiology*. doi: 10.1111/psyp.13650

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:https://doi.org/10.1016/j.brat.2020.103550

Poling, A. (2010). Progressive-ratio schedules and applied behavior analysis. *Journal of Applied Behavior Analysis, 43*, 347-349. doi:10.1901/jaba.2010.43-347

Rodriguez-Romaguera, J., Greenberg, B. D., Rasmussen, S. A., & Quirk, G. J. (2016). An avoidance-based rodent model of exposure with response prevention therapy for obsessive-compulsive disorder. *Biological Psychiatry, 80*, 534-540. doi:10.1016/j.biopsych.2016.02.012

San Martín, C., Jacobs, B., & Vervliet, B. (2020). Further characterization of relief dynamics in the conditioning and generalization of avoidance: Effects of distress tolerance and intolerance of uncertainty. *Behaviour Research and Therapy, 124*, 103526-103526. doi:10.1016/j.brat.2019.103526

Seitz, B. M., Stolyarova, A., & Blaisdell, A. (2019). The modified law of effect explains the partial reinforcement extinction effect. https://doi.org/10.31234/osf.io/kbymr

van den Hout, M. A., Engelhard, I. M., Toffolo, M. B. J., & van Uijen, S. L. (2011). Exposure plus response prevention versus exposure plus safety behaviours in reducing feelings of contamination, fear, danger and disgust. An extended replication of Rachman, Shafran, Radomsky & Zysk (2011). *Journal of Behavior Therapy and Experimental Psychiatry, 42*, 364-370. doi:10.1016/j.jbtep.2011.02.009

Vervliet, B., & Indekeu, E. (2015). Low-cost avoidance behaviors are resistant to fear extinction in humans. *Frontiers in Behavioral Neuroscience, 9*, 351-351. doi:10.3389/fnbeh.2015.00351

Vervliet, B., Lange, I., & Milad, M. R. (2017). Temporal dynamics of relief in avoidance conditioning and fear extinction: Experimental validation and clinical relevance. *Behaviour Research and Therapy, 96*, 66-78. doi:10.1016/j.brat.2017.04.011

Williams, W. A. (2019). Partial reinforcement extinction effect. In Vonk J., and Shackelford T. (eds) *Encyclopedia of Animal Cognition and Behavior*. Springer International Publishing. doi:10.1007/978-3-319-47829-6_809-1

Xia, W., Dymond, S., Lloyd, K., & Vervliet, B. (2017). Partial reinforcement of avoidance and resistance to extinction in humans. *Behaviour Research and Therapy, 96*, 79-89. doi:10.1016/j.brat.2017.04.002

Zuj, D. V., & Norrholm, S. D. (2019). The clinical applications and practical relevance of human conditioning paradigms for posttraumatic stress disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 88*, 339-351. doi:10.1016/j.pnpbp.2018.08.014

Figure Captions

Figure 1. Schematic overview. (A) the experimental design and trial timings of a CS+ trial during *avoidance conditioning* and (B) the sequence of phases and stimulus and response contingencies for all groups. See text for details.

Figure 2. Proportion of avoidance responses for CS+ and CS- for each partial reinforcement group across phases where avoidance was available. Error bars represent SEM.

Figure 3. Trial-by-trial threat expectancy for CS+ (A) and CS- (B) trials per reinforcement group. TC: *threat conditioning*; AV: *avoidance conditioning*; EX: *extinction*; RPE: *response prevention and extinction*; and RE-EX: *re-extinction*. Error bars represent SEM.

Figure 4. Fear ratings for CS+ (A) and CS- (B) trials for each reinforcement group. TC: *threat* conditioning; AV: *avoidance conditioning*; EX: *extinction*; RPE: *response prevention and extinction*; and RE-EX: *re-extinction*. AV_1, AV_2, and so on, refer to sequential blocks of multiple trials binned for analysis purposes. Error bars represent SEM.

Figure 5. Trial-by-trial skin conductance responses (SCRs) for CS+ (A) and CS- (B) trials per reinforcement group. TC: *threat conditioning*; AV: *avoidance conditioning*; EX: *extinction*; RPE: *response prevention and extinction*; and RE-EX: *re-extinction*. Error bars represent SEM.