Human fear conditioning is moderated by stimulus contingency instructions

Gaëtan Mertens[1, 2], Yannick Boddez[3, 4], Angelos-Miltiadis Krypotos[2, 5], and Iris M. Engelhard[2]

[1]Department of Medical and Clinical Psychology, Tilburg University, Tilburg, the Netherlands

[2]Department of Clinical Psychology, Utrecht University, Utrecht, the Netherlands

[3]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

[4]Centre for the Psychology of Learning and Experimental Psychopathology, KU Leuven, Leuven, Belgium

[5]Leuven Research Group on Health Psychology, KU Leuven, Leuven, Belgium

Correspondence concerning this article should be addressed to Gaëtan Mertens, Department of Medical and Clinical Psychology, Warandelaan 2, room T526, Tilburg University, 5037AB Tilburg, the Netherlands.

E-mail: g.mertens@tilburguniversity.edu

Tel: +31134663757

Abstract

Recent research findings indicate that human fear conditioning is affected by instructions, particularly those concerning the contingency between the conditioned stimulus (CS) and the unconditioned stimulus (US). However, whether or not such instructions were provided to participants often remains unsaid in fear conditioning studies. In the current study (N=102), we investigated whether conditioned fear acquisition depends on CS-US contingency instructions. Participants were randomly assigned to one of three groups. The first group was instructed about the precise CS-US contingency before conditioning. The second group was instructed to discover the CS-US contingency. The third group did not receive any contingency instructions. We found facilitated fear acquisition (using skin conductance and startle) and increased contingency awareness in the first and second group compared to the third group. Furthermore, contingency reversal instructions immediately reversed conditioned responses. Based on these results, we advise to systematically report the used contingency instructions in fear conditioning research.

Keywords: Fear Conditioning; Replicability; Instructions; Psychophysiology

1. Introduction

The fear conditioning procedure is a widely used translational paradigm to investigate the etiology and treatment of anxiety-related disorders (Vervliet et al., 2013). In this paradigm, a neutral Conditioned Stimulus (CS) gets paired with an aversive Unconditioned Stimulus (US), which typically results in fearful Conditioned Responses (CRs) towards the CS. It is generally believed that this paradigm models an important etiological pathway for the development of anxiety disorders (De Houwer, 2020; Field, 2006; Mineka & Zinbarg, 2006). Furthermore, it connects fundamental cognitive and psychopharmacological research in animals with clinical research in humans (Haaker et al., 2019).

A growing body of research has indicated that fear conditioning can be influenced by verbal instructions (for a review see: Mertens, Boddez, Sevenster, Engelhard, & De Houwer, 2018). Of particular interest are instructions that are concerned with the contingencies between the CSs and US. Classical (fear) conditioning has been shown to critically rely on contingencies between stimuli (rather than on mere contiguity) (Rescorla, 1988). Furthermore, recent theoretical models have stressed the importance of propositional representations and verbal instructions in human learning (Mitchell et al., 2009). As such, it can be expected that contingency instructions can substantially influence human fear conditioning. Indeed, studies have shown that simply providing participants with the instruction that a CS will be followed by a US is sufficient to install subjective, behavioral, and psychophysiological responses related to fear, without requiring any actual CS-US pairings (e.g., Deltomme, Mertens, Tibboel, & Braem, 2018; Javanbakht et al., 2016; Mertens et al., 2016; Raes, De Houwer, De Schryver, Brass, & Kalisch, 2014).

However, instructions often remain unmentioned in fear conditioning papers. To illustrate, we have analyzed the method sections of 69 empirical articles reporting a fear conditioning study involving human participants published in 2018.[1] Out of the 71 studies reported in these articles, 41 studies (58%) used some type of contingency instructions (precise or general; see below), whereas three studies (4%) explicitly mentioned not using contingency instructions. Crucially, we found that for the remaining 27 (38%) studies, no information was available about the instructions given to the participants. As such, it is unclear whether these 27 studies used contingency instructions or not, because this information is missing. Furthermore, even in those studies that were explicit about the use of instructions, information about the instructions was often minimal (i.e., usually there were no verbatim descriptions of the instructions).

Not reporting the instructions can be problematic for replicating research findings in the fear conditioning literature. For example, verbal instructions can have a considerable impact on the number of participants who show successful fear acquisition. Indeed, several studies have found that, compared to participants who did not receive instructions about the CS-US contingency, participants who did receive such instructions showed stronger differential (i.e., CS+ > CS-) conditioned skin conductance responses (Atlas et al., 2016; Javanbakht et al., 2016; Tabbert et al., 2006), conditioned startle responses (Duits et al., 2017), US expectancy ratings

[1] This was established with a search on PubMed (search syntax: (((fear conditioning) NOT rats) NOT mice) NOT animal). It provided 174 hits, of which 69 articles that included a differential cue fear conditioning procedure with adult human participants were selected for full text screening. We checked whether the articles provided any information about the instructions given to participants prior to the fear conditioning phase (i.e., articles did not necessarily have to state the exact instructions). Forty-two articles provided some information on the instructions given to the participants and 27 articles did not provide any information (for an overview of these studies see https://osf.io/7j56p/).

(Raes et al., 2009), and higher rates of contingency awareness in a post-conditioning assessment (Tabbert et al., 2006). Hence, verbal contingency instructions influence the strength of conditioned fear acquisition. This problem is further exacerbated by the fact that fear conditioning studies commonly exclude participants who do not discriminate between the CSs that are followed by a US from the CSs that are not (sometimes up to 74% of the sample; see Lonsdorf et al., 2019). As such, different contingency instructions can result in a different sample selection, which can further complicate the replication of prior research findings.

The contingency instructions provided at the start of a conditioning procedure can be generally divided into three types. First, participants can be informed about the precise contingency between the CS and US (e.g., Atlas et al., 2016; Bublatzky, Gerdes, & Alpers, 2014; Costa, Bradley, & Lang, 2015; Mertens & De Houwer, 2016). With such instructions, participants know, before any stimulus pairings, which CS predict the US. This is also sometimes referred to as *instructed fear conditioning* (because participants learn about CS-US contingencies). Second, participants can be informed that there is a contingency in the task and be encouraged to discover this contingency, but without being told which of the CSs will be followed by the US (Engelhard et al., 2015; Golkar et al., 2012; Haesen & Vervliet, 2015; Mertens et al., 2019). With these instructions, participants know that there is a CS-US contingency, but they do not know the exact CS-US relationship. This requires them to learn the CS-US contingencies on the basis of CS-US pairings, as in a typical (uninstructed) fear conditioning procedure. It most likely focuses participants' attention to the contingencies, which will facilitate learning (Mackintosh, 1975). Finally, participants sometimes do not receive any information about the contingencies in the task, but are only told that different stimuli will be presented (i.e., *uninstructed fear conditioning*; e.g., Haaker et al., 2015; Leuchs, Schneider, &

Spoormaker, 2018; Miskovic & Keil, 2013; Sjouwerman, Niehaus, Kuhn, & Lonsdorf, 2016). Without contingency instructions, participants have no prior knowledge about the presence of contingencies in the conditioning procedure and can only rely on the actual stimulus pairings to learn the CS-US contingencies. To our knowledge, no studies so far have directly compared the effects of different types of CS-US contingency instructions on conditioned fear acquisition.

Participants can also be given instructions before other phases of fear conditioning, such as fear extinction (i.e., when a CS is no longer followed by a US), generalization (generalization of CRs to other CSs), and return of fear (return of CRs after fear extinction) (Lonsdorf et al., 2017). For example, in research from our group and other research groups investigating contextual renewal (i.e., the return of conditioned fear due to a change in contextual features) in a 2-day paradigm, participants were instructed to "think back to what you learned the previous day" on the second day (Landkroon et al., 2019; Milad et al., 2005). Such instructions may affect contextual renewal, by hinting that what was learned previously is still relevant. However, relatively few studies so far have tested whether contingency instructions affect conditioned responses beyond the acquisition phase (for work on instructed extinction and instructed reversal learning see Luck & Lipp, 2016b and Mertens et al., 2018).

In the current study, we aimed to assess the effect of these three different types of contingency instructions (i.e., precise contingency instructions, general contingency instructions, and no contingency instructions) on conditioned fear acquisition. Particularly, fear acquisition was measured with psychophysiological discrimination between the CS+ (i.e., the CS paired with the US) and the CS- (i.e., the CS not paired with the US) at the end of the acquisition phase, and contingency awareness rates as assessed with a retrospective questionnaire. We expected that participants in the *precise contingency instructions* and *general contingency instructions*

conditions would show higher fear acquisition rates and contingency awareness than participants in the no contingency instructions condition. Furthermore, we expected only a slight advantage in the precise contingency instructions condition compared to the general contingency instructions condition, because most participants in the general contingency instructions condition are expected to discover the contingencies as well (see Section 2.2).

An additional aim of this study was to replicate prior studies showing that contingency instructions can also influence conditioned responses later on in a conditioning procedure. Given recent concerns with regard to the replication of findings within psychology (Klein et al., 2014; Open Science Collaboration, 2015), we found it important to provide a replication of such findings. In addition, we want to provide an extension of those findings to fear potentiated startle, which has been a relatively rarely used measure in this type of studies (Luck & Lipp, 2015; Mertens & De Houwer, 2016). For this purpose, we provided participants with contingency reversal instructions after the acquisition phase. As observed in previous studies (Atlas et al., 2016; Luck & Lipp, 2016a; Mertens & De Houwer, 2016; Morriss et al., 2019; Wilson, 1968), we expected that these contingency instructions would reverse conditioned psychophysiological responses.

2. Methods

2.1. Preregistration and data availability

This study was preregistered on the Open Science Framework at the following link: https://doi.org/10.17605/OSF.IO/7J56P. Raw and final datafiles can be obtained through this link as well. The procedure of this study falls within a research line of fear conditioning studies, which has already received approval by the ethics committee of the Faculty of Social and Behavioral Science at Utrecht University (FETC16-054).

2.2. Participants

One hundred and two students from Utrecht University were recruited to participate in this study. Mostly English speaking international undergraduate students were recruited to reduce overlap with the target population of other ongoing (Dutch) fear conditioning studies from our research group. Participants were randomly allocated to one of the three different conditions in the experiment (i.e., precise contingency instructions condition, general contingency instructions condition, and no contingency instructions condition; n = 34 per condition). The sample size was determined using an a-priori power analysis. Particularly, under the assumption that 90% of the participants in the precise contingency instructions condition, 80% in the general contingency instructions condition, and 50% in the no contingency instructions condition show successful fear acquisition ($\omega = 0.384$)[2], a total sample of 102 participants was required to detect a significant effect ($p = .025$; see below) with a power of > 0.9 (Faul et al., 2007). The data of six participants were removed due to technical reasons (see Section 2.5.3.1) and were replaced by the data of six new participants. Participants were recruited through flyers and posters on campus and were screened for self-reported physical and mental health. All participants completed an informed consent form and were instructed that they could discontinue the experiment at any point without any negative consequences. Participants received financial compensation (€8) or course credit in exchange for their participation. Table 1

[2] These numbers were our "best guess" based on our experience with fear conditioning procedures, given that there are no reliable estimates in the literature regarding the rates of successful fear acquisition rates in the different contingency instructions conditions. Note nonetheless that exclusion rates of 50% and more have been reported in the literature (Lonsdorf et al., 2019; Nitta et al., 2020; Reddan et al., 2018). As such, we used this as an estimate for the no contingency instructions condition. Furthermore, several studies have shown facilitated fear acquisition due to contingency instructions (Javanbakht et al., 2016; Mertens et al., 2016; Raes et al., 2014). As such, we estimated higher rates of fear acquisition in the contingency instructions conditions, with the highest rate in the specific contingency instructions condition because of the more detailed instructions.

provides more detailed demographic information, trait anxiety scores, selected US intensity, and US pain ratings regarding the participants in the three different conditions of this study (see below).

-INSERT TABLE 1 HERE-

2.3. Materials

2.3.1. Apparatus. The experiment was programmed in Inquisit (v4) and run on a HP Z230 desktop computer running Windows 8.1 Pro. The electrical simulation was generated with a Digitimer DS7A system. Skin conductance was measured using a Biosemi bio-amplifier and two Biosemi GSR electrodes filled with Signa electrode gel attached to the thenar and hypothenar of the left hand. Startle responses were measured with two BioSemi EMG electrodes attached below the left eye (Blumenthal et al., 2005). Psychophysiological measures were collected with Actiview and further analyzed offline with BrainVision Analyzer 2.0 software.

2.3.2. Questionnaires. Trait anxiety was determined with the State-Trait Anxiety Inventory – trait version (STAI-T, range: 20-80; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983; van der Ploeg, Defares, & Spielberger, 2000) to control for possible differences in trait anxiety between the conditions. Additionally, the short version of the Intolerance of Uncertainty Scale (Carleton et al., 2007) and the Context Sensitivity Index (Bonanno et al., 2020) were completed by the participants for unrelated research questions, and will therefore not be reported here.

2.3.3. Stimuli. The unconditioned stimulus (US) was a 558-ms electrical stimulation (i.e., nine pulses of 2 ms with a 60 ms inter-pulse interval) administered through two electrodes attached to the wrist of the right hand. The intensity of this stimulus was individually set for each participant with a work-up procedure (see the Procedure section).

Conditioned stimuli (CSs) were two grey geometrical shapes (a circle and a square), which are common stimuli in fear conditioning procedures (Lonsdorf et al., 2017). These were presented with a resolution of 300 by 300 pixels on a 23-inch screen (screen resolution: 1920 by 1080 pixels).

2.4. Procedure

2.4.1. Startup and work-up procedure. Upon arrival in the lab, participants washed their hands and were then asked to read the information letter about the experiment, provide informed consent and complete the STAI-T. Next, skin conductance and electrical stimulation electrodes were attached. Participants went through a work-up procedure in which the US intensity was determined. They were asked to select an intensity level that was unpleasant but tolerable (Mertens & De Houwer, 2016). To operationalize the intensity, participants were asked to score the intensity of the US on a 0 to 10 scale (0 = *no pain at all*, 10 = *maximum level to voluntarily tolerate*). The work-up procedure stopped when participants rated the intensity as 7 or higher. The final intensity level was used in the experiment, unless participants indicated before reaching 7 that they did not want to increase the intensity further. In these latter cases, the maximal tolerable intensity was used (Mertens & De Houwer, 2016).

2.4.2. Instruction manipulation. After the work-up procedure, participants were randomly allocated to one of the three conditions. Particularly, in the *precise contingency*

instructions condition, they received the following instructions on the computer screen: "*In the following experiment you will see two different shapes appear on the screen: A square and a circle. The square*[/circle] *will sometimes be followed by an electrical shock and the circle*[/square] *will never be followed by an electrical shock.*" Participants in the *general contingency instructions* condition received these instructions: "*In the following experiment you will see two different shapes appear on the screen: A square and a circle. One of the shapes will sometimes be followed by an electrical shock and the other shape will never be followed by an electrical shock. Your task is to learn to predict when the shock will be presented.*" Finally, participants in the *no contingency instructions* condition received the following instructions: "*In the following experiment you will see two different shapes appear on the screen: A square and a circle. You will also sometimes receive an electrical shock.*" Following the instruction manipulation, participants were told to press the spacebar to continue with the experiment.

2.4.3. Startle habituation and fear conditioning phase. Subsequently, participants habituated to the startle probe (50 ms, 95 dB). They heard the probe 10 times with a 7 s inter-trial interval (ITI), which was immediately followed by the fear conditioning phase. This phase consisted of eight presentations of the circle and the square. Counterbalanced, either the circle or square was followed by the electrical stimulation on six out of the eight trials (75% reinforcement rate). Each shape was shown for 8 s. In each trial, a startle probe was presented 7 s after CS onset. In case of a reinforced trial, the US was administered immediately at CS offset. The ITI was either 12, 14 or 16 s. The order of CS presentations was semi-random with the restriction of maximally two identical consecutive trials.

2.4.4. Questions regarding contingency awareness. After the acquisition phase, participants were asked about their awareness of the CS-US contingencies with the following two questions about each CS: (1) "Was the square[/circle] followed by the electric shock?", response options: "Yes", "No"; and (2) "How certain are you about your answer?"; response options: "very sure", "quite sure", "quite unsure", and "very unsure". A comparable procedure of assessing contingency awareness has been used in other fear conditioning studies (e.g., Singh et al., 2013; Tabbert et al., 2006; Wegerer, Blechert, Kerschbaum, & Wilhelm, 2013). It should be noted though that, despite its common use as a way of establishing contingency awareness, retrospective questionnaires remain an imperfect and insensitive measure of contingency awareness (Dawson & Reardon, 1973; Mertens & Engelhard, 2020).

2.4.5. Contingency reversal instructions and reversal phase. Following the acquisition phase and contingency awareness assessment, participants were instructed that stimulus contingencies would be reversed in the following phase ("*In the next phase of the experiment, the relationship between the shapes and the electric shock will be reversed: The square[/circle] WILL now NOT be followed by the electric shock. The circle[/square] WILL now SOMETIMES be followed by the electric shock.*"). This instruction was identical for participants in all the conditions. Following these instructions, the experiment continued with the same procedure as in the acquisition phase, except that CS+ and CS- were each shown five times (instead of eight) and the CS+ was only reinforced once after the third trial. We decided to only reinforce the CS+ after three trials to ensure that reversal up to this point was only based on the verbal contingency instructions. After the reversal phase, participants were asked to indicate the contingencies of the previous phase. Finally, they were debriefed and compensated for their participation.

2.5. Data Preprocessing and Analysis

2.5.1. Skin conductance responses (SCRs). SCRs were calculated by subtracting the mean value of a baseline period (2 s before CS onset) from the highest peak during the 1 to 8 s interval post CS onset (Pineles et al., 2009). Thereafter, skin conductance values were range corrected using the largest response for each participant and square root transformed to normalize the data (Dawson et al., 2007). A minimum response criterion was set at .02 μS.

2.5.2. Fear potentiated startle (FPS). The electromyography signal of the startle response was filtered (28-500 Hz), smoothed (15.9 Hz low-pass filter), and rectified. Startle magnitude was calculated by subtracting the baseline value (time window: 0-20 ms after probe onset) from the highest peak value in the 21 to 150 ms time window after startle probe onset. These values were then T-transformed using each participants' individual mean and standard deviation (Blumenthal et al., 2005).

2.5.3. Data analysis.

2.5.3.1. Data exclusion. The data of six participants was excluded due to incorrect storage of the data (n = 4) or problems with the storage of markers in the datafiles (n = 2). These data were replaced with data of six new participants to maintain our targeted sample size (n = 102; see the Participants section).

2.5.3.2. Planned statistical analyses. The focus of this study was on the number of participants who show successful conditioned fear acquisition. This was defined as a positive difference between the CS+ and CS- at the end of the acquisition phase (i.e., Fear_CS+ - Fear_CS- > 0; see the preregistration file). To reduce the influence of error variance, we

averaged responses of the last two acquisition trials to calculate this index. This criterion is commonly used for fear acquisition (e.g., Ahmed & Lovibond, 2015; Atlas et al., 2016; Golkar, Tjaden, & Kindt, 2017; Javanbakht et al., 2016; Klucken et al., 2016; Morriss, Christakou, & van Reekum, 2016) and is straightforward to interpret and implement. Participants who did not meet this criterion were coded as unsuccessful fear acquisition. We tested whether the rate of participants who showed successful fear acquisition differed between the different conditions by conducting a Chi-square test. Because the same focal hypothesis was tested both with SCRs and FPS, an alpha-value of .025 (i.e., 0.5/2) was used (see our preregistration).

In addition to successful acquisition of conditioned fear, we investigated the number of participants who showed successful contingency awareness. Contingency awareness is a common exclusion criterion in fear conditioning research (e.g., Dirikx, Vansteenwegen, Eelen, & Hermans, 2009; Golkar et al., 2017; Mertens et al., 2019; Rowles, Lipp, & Mallan, 2012). We considered participants to be contingency aware if they correctly indicated which CS was followed by the electrical stimulation and which CS was not followed by the electrical stimulation (see Singh et al., 2013). Furthermore, participants had to be “completely certain” or “fairly certain” of their answer for both the CS+ and CS- to account for guessing. Otherwise, they were categorized as contingency unaware. The number of contingency aware participants in the three contingency instructions conditions was analyzed using a Chi-square test.

2.5.3.3. Secondary analyses. In additional and preregistered secondary analyses (see preregistration file), results of SCR and FPS in the acquisition phase were analyzed using a repeated measures ANOVA with factors instructions (between-subjects: precise contingency, general contingency and no contingency), CS type (within-subjects: CS+ and CS-) and trial

(within-subjects: 1 to 8). This analysis takes into account all the trials of the acquisition phase, rather than only the last two trials, which allows us to also test the course of learning rather than only the final two trials. Furthermore, the continuous nature of this analysis provides more statistical power than the primary analysis.

Furthermore, we analyzed results of the reversal phase by comparing the average of the two last trials of the acquisition phase to the average of the two first trials in the reversal phase using a repeated measures ANOVA with factors instructions (between-subjects: precise contingency, general contingency and no contingency), CS type (within-subjects: CS+ and CS-) and phase (within-subjects: acquisition and reversal).

All analyses were run in SPSS v25, using an alpha-value cut-off of .05, unless otherwise stated. Violations of the sphericity assumption were corrected using Greenhouse-Geisser corrections.

3. Results

3.1. Primary analyses

3.1.1. Successful psychophysiological fear acquisition. Table 2 provides the results regarding successful CS discrimination for both SCR and FPS. Descriptively, successful acquisition rates were higher in the precise contingency instructions and general contingency instructions conditions for SCR and in the general contingency instructions condition for FPS. Statistically, however, we did not find evidence for differences in acquisition rates between the conditions (*p*-values > .05; see Table 2).

3.1.2. Contingency awareness rates. Table 2 also provides the results for retrospective contingency awareness rates in the different conditions. As expected, the percentage of contingency aware participants after the acquisition phase was significantly higher in the general contingency instructions condition than in the no contingency instructions condition ($\chi^2(1) = 11.88$, $p = .001$). However, unexpectedly, it was not significantly higher in the precise contingency instructions condition than in the no contingency instructions condition ($\chi^2(1) = 2.51$, $p = .113$). Also surprisingly, the percentage of contingency aware participants in the general contingency instructions condition was significantly higher than in the precise contingency instructions condition ($\chi^2(1) = 4.17$, $p = .041$).

With regard to contingency awareness rates after the reversal phase, there were no significant differences between the different conditions (all $\chi^2(1) < 1.5$, $p > .25$). The relative low awareness rates after the reversal phase were due to a large proportion of participants reporting low confidence in their responses. That is, 39 (38%) of the participants indicated being "quite unsure" or "very unsure" about their responses after the reversal phase (compared to 18% after the acquisition phase), which resulted in them being classified as contingency unaware (regardless of the correctness of their answer; see Section 2.5.3.2). This drop in confidence ratings was likely due to the low reinforcement rate in the reversal phase (i.e., 20%).

-INSERT TABLE 2 HERE-

3.2. Secondary analyses and effects of reversal instructions

3.2.1. Skin conductance responses.

3.2.1.1. Acquisition phase. The repeated measures ANOVA of the results of the acquisition phase revealed main effects of CS type, $F(1, 99) = 92.00$, $p < .001$, $\eta^2_p = 0.48$, and trial, $F(6.02, 595.67) = 13.20$, $p < .001$, $\eta^2_p = 0.12$. These main effects were qualified by an interaction between CS type and trial, $F(7, 693) = 2.48$, $p = .016$, $\eta^2_p = 0.02$, and, crucially, between CS type and condition, $F(2, 99) = 5.52$, $p = .005$, $\eta^2_p = 0.10$. The interaction between CS type and trial was due to gradual acquisition of differential conditioning, as indicated by the significant differentiation between CS+ and CS- on the last trial of the acquisition phase, $t(101) = 5.10$, $p < .001$, while responding to CS+ and CS- did not differ on the first trial of the acquisition phase, $t(101) = 1.16$, $p = .251$ (see Figure 1). The interaction between CS type and condition was followed up by direct comparisons between the different conditions. These indicated stronger differentiation between CS+ and CS- in the precise contingency condition compared to the no contingency condition, CS type and condition interaction: $F(1, 67) = 8.88$, $p = .004$, $\eta^2_p = 0.12$, and in the general contingency condition compared to the no contingency condition, CS type and condition interaction: $F(1, 65) = 7.01$, $p = .010$, $\eta^2_p = 0.10$. The precise and general contingency conditions did not differ significantly from one another, CS type and condition interaction: $F(1, 66) = 0.21$, $p = .649$, $\eta^2_p < 0.01$ (see Figure 1). No other main or interaction effects were significant, F-values < 1.5, p-values > .15, $\eta^2_p < 0.03$.

3.2.1.2. Reversal phase. The repeated measures ANOVA assessing the effect of the reversal instructions indicated main effects of CS type, $F(1, 99) = 7.18$, $p = .009$, $\eta^2_p = 0.07$, and phase, $F(1, 99) = 7.30$, $p = .008$, $\eta^2_p = 0.07$, and, crucially, an interaction effect between CS type and phase, $F(1, 99) = 38.78$, $p < .001$, $\eta^2_p = 0.28$. This interaction effect was due to lower CS+

SCR values after the reversal instruction (M = 0.32, SD = 0.25) compared to before (M = 0.40, SD = 0.30), $t(101) = 2.47$, $p = .015$, whereas this pattern was the reverse for CS- (before reversal instructions: M = 0.20, SD = 0.23; after reversal instructions: M = 0.41, SD = 0.30; see Figure 1), $t(101) = -6.77$, $p < .001$. Significant differential conditioning (CS+ > CS-) was observed at the end of the acquisition phase, $t(101) = 7.44$, $p < .001$, and significant reversal (CS- > CS+) was obtained after the reversal instructions, $t(101) = -2.43$, $p = .017$. The other main and interaction effects were not significant, F-values < 2.1, p-values > .13, $\eta^2_p < 0.05$.

-INSERT FIGURE 1 HERE-

3.2.2. Fear potentiated startle.

3.2.2.1. Acquisition. Similar to the results of SCR, the repeated measures ANOVA of the FPS data during the acquisition phase revealed main effects of CS type, $F(1, 99) = 47.84$, $p < .001$, $\eta^2_p = 0.33$, and trial, $F(5.91, 585.09) = 12.42$, $p < .001$, $\eta^2_p = 0.11$. These main effects were qualified by a three-way interaction between CS type, trial, and condition, $F(14, 693) = 1.74$, $p = .044$, $\eta^2_p = 0.03$. Breaking down this interaction, with separated CS by trial repeated measures ANOVA's, only a clear interaction between CS type and trial was observed in the general contingency instruction condition, $F(7, 224) = 2.95$, $p = .006$, $\eta^2_p = 0.08$, whereas no such interaction was observed for the precise contingency instruction condition, $F(7, 238) = 0.56$, $p = .785$, $\eta^2_p = 0.02$, or the no contingency instruction condition, $F(7, 231) = 0.74$, $p = .640$, $\eta^2_p = 0.02$. The significant interaction between CS type by trial in the general contingency

instructions was due to gradual acquisition of differential conditioning, as evidenced by significant differential conditioning on the last trial of the acquisition phase, $t(32) = 2.06$, $p = .048$, while there was no difference between the CS+ and CS- on the first trial of the acquisition phase, $t(32) = -0.28$, $p = .781$ (see Figure 2). In all conditions, a significant effect of CS type was observed (p-values < .017), but the effect was more pronounced in the precise contingency ($\eta^2_p = 0.39$) and the general contingency instruction conditions ($\eta^2_p = 0.42$), than in the no contingency instruction condition ($\eta^2_p = 0.16$) (see Figure 2). No other main or interaction effects were significant, F-values < 1.7, p-values > .2, $\eta^2_p < 0.04$.

3.2.2.2. Reversal. As for the SCR results, the repeated measures ANOVA assessing the effect of the reversal instructions for FPS indicated a main effect of phase, $F(1, 99) = 5.72$, $p = .019$, $\eta^2_p = 0.06$, an interaction effect between instructions and phase, $F(2, 99) = 3.38$, $p = .038$, $\eta^2_p = 0.06$, and, crucially, an interaction effect between CS type and phase, $F(1, 99) = 53.98$, $p < .001$, $\eta^2_p = 0.35$. The latter interaction effect was due to lower CS+ FPS values after the reversal instruction (M = 47.04, SD = 6.27) compared to before (M = 50.32, SD = 7.74), $t(101) = 3.27$, $p = .001$, whereas this pattern was the reverse for CS- (before reversal instructions: M = 45.82, SD = 5.42; after reversal instructions: M = 52.23, SD = 7.17; see Figure 2), $t(101) = -7.19$, $p < .001$. Significant differential conditioning (CS+ > CS-) was observed at the end of the acquisition phase, $t(101) = 5.12$, $p < .001$, and significant reversal (CS- > CS+) was obtained after the reversal instructions, $t(101) = -5.49$, $p < .001$. The other main and interaction effects were not significant, F-values < 2.2, p-values > .12, $\eta^2_p < 0.05$.

-INSERT FIGURE 2 HERE-

4. Discussion

We investigated the effects of contingency instructions prior to conditioning and contingency reversal instructions after conditioning on conditioned psychophysiological responses (SCR and FPS). Contingency instructions before an acquisition phase affected the rate of contingency aware participants, particularly when participants were instructed to discover the contingencies themselves. Using a dichotomic criterion of successful conditioning, we did not observe significant effects of contingency instructions on successful fear acquisition rates for SCR and FPS, although numerically the results were in the expected direction. Secondary, but also pre-registered, continuous analyses on the trial-by-trial data in the acquisition phase did, however, reveal a significant effect of contingency instructions for SCR and FPS, indicating that precise and general contingency instructions resulted in more pronounced differential conditioning compared to no contingency instructions (see Figures 1 and 2). Finally, reversal instructions following the fear acquisition phase reversed conditioned responses with both SCR and FPS. Collectively, these results provide empirical evidence that contingency instructions influence conditioned fear responses.

It is important to note that we pre-registered primary and secondary analyses. In the primary analyses, we focused on dichotomous measures of successful fear acquisition and contingency awareness, because they are commonly used as exclusion criteria in fear conditioning research (e.g., Ahmed & Lovibond, 2015; Atlas et al., 2016; Golkar, Tjaden, & Kindt, 2017; Javanbakht et al., 2016; Klucken et al., 2016; Lonsdorf et al., 2019; Morriss et al., 2016). As such, these dichotomous measures influence the selection of the final sample in a

substantial portion of fear conditioning studies, even though it was unclear to what extent these measures are affected by contingency instructions (i.e., given that the instructions are not always reported in fear conditioning papers; see the Introduction). However, arguments against this analysis are that it lacks sensitivity due to not taking into account the trial-by-trial variability in physiological responses and due to the dichotomous nature of the outcomes. Therefore, we also registered the secondary approach for analyzing our data, which involved analyzing the physiological responses in a continuous and trial-by-trial fashion. This provided a more sensitive test for our hypothesis. Furthermore, it is important to note that there are typically multiple valid ways to analyze a single dataset and these do not always provide identical results (e.g., Silberzahn et al., 2018). As such, we decided that it is most transparent to report the results of both our registered data analysis approaches.

We think that our results have important implications for the human fear conditioning field. First, researchers should clearly indicate which instructions they gave participants in all phases of the experiment either in the main paper or in a supplemental file to the main paper. Such information is crucial for a full evaluation of the results and the replication of published research. Currently, a substantial part of the literature fails to report this important methodological aspect (i.e., 38% in 2018, see the Introduction). Such unreported variation in methodological details can complicate the interpretation and replication of research findings, as the results may hinge on the verbal instructions provided to participants.

A second implication of our results is that verbal instructions should be considered in the design of fear conditioning studies. Precise or general verbal contingency instructions can strengthen both psychophysiological conditioning and contingency awareness rates. Hence, such

instructions could be used to obtain more robust conditioning. Notably, this has been done in studies that rely on fear acquisition to examine individual differences in extinction learning or interventions to target acquired fear (Leer et al., 2013; Lommen et al., 2013). However, it should be noted that it is not yet clear how verbal contingency instructions interact with other phenomena in which researchers are often interested, such as extinction and return of fear. This should be addressed in future studies. In addition, some research questions may necessitate uninstructed learning of the contingencies. In this case, attention should be devoted to avoiding any references in the instructions to the contingencies in the task, as this may affect the spontaneous learning of the contingencies. Importantly, given the common practice of excluding participants based on unsuccessful fear acquisition (Chalkia et al., 2020; Lonsdorf et al., 2019) or lack of contingency awareness (e.g., Mertens et al., 2019; Wegerer et al., 2013), verbal contingency instructions can affect the final constellation (and hence statistical power and representativeness) of the sample and should therefore be reported. Finally, not only in the initial fear acquisition phase, but also in subsequent phases of conditioning experiment (e.g., extinction phase, generalization phase, return of fear), the effects of verbal instructions should be considered. That is, references to the contingencies in the instructions (e.g., “think back to the contingencies in the previous phase”) may influence the results in this phase as well (e.g., stronger return of fear). Hence, researchers should also clarify in their papers which instructions were given in other phases of fear conditioning experiments.

With regard to theoretical implications, the results of this study provide further support for models that assign a role to inferential reasoning processes in human learning. There is ongoing debate about the processes that underlie human (fear) conditioning. Some authors have

proposed that human fear conditioning, and in particular conditioning of psychophysiological measures, occurs largely automatically (i.e., without effort and outside of voluntary control) and without awareness (LeDoux & Pine, 2016). These theorists have also argued that verbal instructions would only have a minimal impact in fear conditioning (Olsson & Phelps, 2007). In contrast, others have argued that human (fear) conditioning requires inferential reasoning and is sensitive to verbal instructions (Lovibond, 2011; Mertens & Engelhard, 2020; Mitchell et al., 2009). Our results lend more support to the latter class of models by demonstrating the clear impact of verbal contingency instructions on fear conditioning, including psychophysiological measures of fear. Nonetheless, it should be noted that sensitivity to verbal instructions does not provide direct evidence for the involvement of inferential reasoning in classical conditioning. Establishing the involvement of inferential processes in fear conditioning requires manipulation of this process (e.g., by influencing the inferences that participants draw; for such demonstrations see Lovibond, 2003; Raes et al., 2011).

A number of relevant limitations of this work can be noted. First, the use of a retrospective questionnaire to assess contingency awareness is considered to be a valid (Dawson & Reardon, 1973), though insensitive test (Lovibond & Shanks, 2002; Mertens & Engelhard, 2020). A more sensitive approach would be to measure contingency awareness in a trial-by-trial fashion on a continuous scale (Lovibond & Shanks, 2002). Nonetheless, retrospective questionnaires are still often used in the literature and therefore our results provide relevant information on how contingency instructions can influence a regularly used measure for contingency awareness (e.g., Singh et al., 2013; Tabbert et al., 2006; Wegerer et al., 2013). Second, we did not measure all types of conditioned responses that can be collected in fear

conditioning research such as self-reported US expectancy or valence ratings, avoidance behaviors, or other types of psychophysiological responses such a heart rate or functional brain imaging. Despite not providing direct evidence for the effects of instructions on these other types of conditioned responses, there are no a priori reasons to presume that our results and recommendations are not relevant for these outcome measures as well. Indeed, there is already evidence that instructions can influence these other types of conditioned responses (see Mertens et al., 2018). Another limitation is that we only considered the effects of instructions in one specific version of the fear conditioning paradigm (i.e., using geometrical shapes as CSs and an electrical shock as the US, without trial-by-trial subjective ratings, in a healthy student sample, and using a 75% reinforcement schedule). It is conceivable that the effects of verbal contingency instructions are more or less outspoken when using different parameters (e.g., using 100% reinforcement) or relying on different populations. The interaction between such parameters and the effects of verbal instructions needs to be further investigated. A third limitation was the inclusion of startle probes. This can interfere with the acquisition of conditioned SCRs and contingency awareness, which complicates the interpretation of the results for these measures (Sjouwerman et al., 2016). Nonetheless, clear differential fear acquisition was observed for SCRs and contingency awareness rates were quite high (particularly in the general contingency instructions condition), suggesting that the inclusion of startle probes did not strongly affect learning as indexed by these measures.

In conclusion, the results of this study highlight that human fear conditioning can be substantially influenced by verbal instructions provided to participants. As such, fear conditioning researchers are advised to report the instructions that were given to participants.

References

Ahmed, O., & Lovibond, P. F. (2015). The Impact of Instructions on Generalization of Conditioned Fear in Humans. *Behavior Therapy*, *46*(5), 597–603. https://doi.org/10.1016/j.beth.2014.12.007

Atlas, L. Y., Doll, B. B., Li, J., Daw, N. D., & Phelps, E. A. (2016). Instructed knowledge shapes feedback-driven aversive learning in striatum and orbitofrontal cortex, but not the amygdala. *ELife*, *5*(MAY2016), 1–26. https://doi.org/10.7554/eLife.15192

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V, & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*(1), 1–15. https://doi.org/10.1111/j.1469-8986.2005.00271.x

Bonanno, G. A., Maccallum, F., Malgaroli, M., & Hou, W. K. (2020). The Context Sensitivity Index (CSI): Measuring the Ability to Identify the Presence and Absence of Stressor Context Cues. *Assessment*, *27*(2), 261–273. https://doi.org/10.1177/1073191118820131

Bublatzky, F., Gerdes, A. B. M., & Alpers, G. W. (2014). The persistence of socially instructed threat: Two threat-of-shock studies. *Psychophysiology*. https://doi.org/10.1111/psyp.12251

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

Chalkia, A., Van Oudenhove, L., & Beckers, T. (2020). Preventing the return of fear in humans using reconsolidation update mechanisms: A verification report of Schiller et al. (2010). *Cortex*, *129*, 510–525. https://doi.org/10.1016/j.cortex.2020.03.031

Costa, V. D., Bradley, M. M., & Lang, P. J. (2015). From threat to safety: Instructed reversal of defensive reactions. *Psychophysiology*, *52*(3), 325–332. https://doi.org/10.1111/psyp.12359

Dawson, M. E., & Reardon, P. (1973). Construct validity of recall and recognition postconditioning measures of awareness. *Journal of Experimental Psychology*, *98*(2), 308–315. http://www.ncbi.nlm.nih.gov/pubmed/4705630

Dawson, M. E., Schell, A. M., Filion, D. L., & Berntson, G. G. (2007). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology* (3rd ed., pp. 157–181). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

De Houwer, J. (2020). Revisiting classical conditioning as a model for anxiety disorders: A conceptual analysis and brief review. *Behaviour Research and Therapy*, *127*, 103558. https://doi.org/10.1016/j.brat.2020.103558

Deltomme, B., Mertens, G., Tibboel, H., & Braem, S. (2017). Instructed fear stimuli bias visual attention. *Acta Psychologica*. https://doi.org/10.1016/j.actpsy.2017.08.010

Dirikx, T., Vansteenwegen, D., Eelen, P., & Hermans, D. (2009). Non-differential return of fear in humans after a reinstatement procedure. *Acta Psychologica*, *130*(3), 175–182. https://doi.org/10.1016/j.actpsy.2008.12.002

Duits, P., Richter, J., Baas, J. M. P., Engelhard, I. M., Limberg-Thiesen, A., Heitland, I., Hamm, A. O., & Cath, D. C. (2017). Enhancing Effects of Contingency Instructions on Fear Acquisition and Extinction in Anxiety Disorders. *Journal of Abnormal Psychology*. https://doi.org/10.1037/abn0000266

Engelhard, I. M., van Uijen, S. L., van Seters, N., & Velu, N. (2015). The Effects of Safety Behavior Directed Towards a Safety Cue on Perceptions of Threat. *Behavior Therapy*, *46*(5), 604–610. https://doi.org/10.1016/j.beth.2014.12.006

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. https://doi.org/10.3758/BF03193146

Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias? *Clinical Psychology Review*, *26*(7), 857–875. https://doi.org/10.1016/j.cpr.2005.05.010

Golkar, A., Bellander, M., Olsson, A., & Ohman, A. (2012). Are fear memories erasable?-reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli. *Frontiers in Behavioral Neuroscience*, *6*(November), 80. https://doi.org/10.3389/fnbeh.2012.00080

Golkar, A., Tjaden, C., & Kindt, M. (2017). Vicarious extinction learning during reconsolidation neutralizes fear memory. *Behaviour Research and Therapy*, *92*, 87–93. https://doi.org/10.1016/j.brat.2017.02.004

Haaker, J., Lonsdorf, T. B., Schümann, D., Menz, M., Brassen, S., Bunzeck, N., Gamer, M., & Kalisch, R. (2015). Deficient inhibitory processing in trait anxiety: Evidence from context-dependent fear learning, extinction recall and renewal. *Biological Psychology*, *111*(day 2), 65–72. https://doi.org/10.1016/j.biopsycho.2015.07.010

Haaker, J., Maren, S., Andreatta, M., Merz, C. J., Richter, J., Richter, S. H., Meir Drexler, S., Lange, M. D., Jüngling, K., Nees, F., Seidenbecher, T., Fullana, M. A., Wotjak, C. T., & Lonsdorf, T. B. (2019). Making translation work: Harmonizing cross-species methodology in the behavioural neuroscience of Pavlovian fear conditioning. *Neuroscience & Biobehavioral Reviews*, *107*(April), 329–345. https://doi.org/10.1016/j.neubiorev.2019.09.020

Haesen, K., & Vervliet, B. (2015). Beyond extinction: Habituation eliminates conditioned skin conductance across contexts. *International Journal of Psychophysiology*, *98*(3), 529–534.

https://doi.org/10.1016/j.ijpsycho.2014.11.010

Javanbakht, A., Duval, E. R., Cisneros, M. E., Taylor, S. F., Kessler, D., & Liberzon, I. (2016). Instructed fear learning, extinction, and recall: additive effects of cognitive information on emotional learning of fear. *Cognition and Emotion*, *9931*(May), 1–8. https://doi.org/10.1080/02699931.2016.1169997

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating Variation in Replicability. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klucken, T., Kruse, O., Schweckendiek, J., Kuepper, Y., Mueller, E. M., Hennig, J., & Stark, R. (2016). No evidence for blocking the return of fear by disrupting reconsolidation prior to extinction learning. *Cortex*, *79*, 112–122. https://doi.org/10.1016/j.cortex.2016.03.015

Landkroon, E., Mertens, G., Sevenster, D., Dibbets, P., & Engelhard, I. M. (2019). Renewal of conditioned fear responses using a film clip as the aversive unconditioned stimulus. *Journal of Behavior Therapy and Experimental Psychiatry*, *65*, 101493. https://doi.org/10.1016/j.jbtep.2019.101493

LeDoux, J. E., & Pine, D. S. (2016). Using Neuroscience to Help Understand Fear and Anxiety: A Two-System Framework. *American Journal of Psychiatry*, *173*(11), 1083–1093. https://doi.org/10.1176/appi.ajp.2016.16030353

Leer, A., Engelhard, I. M., Altink, A., & van den Hout, M. A. (2013). Eye movements during recall of aversive memory decreases conditioned fear. *Behaviour Research and Therapy*, *51*(10), 633–640. https://doi.org/10.1016/j.brat.2013.07.004

Leuchs, L., Schneider, M., & Spoormaker, V. I. (2018). Measuring the conditioned response: A comparison of pupillometry, skin conductance, and startle electromyography. *Psychophysiology*, e13283. https://doi.org/10.1111/psyp.13283

Lommen, M. J. J., Engelhard, I. M., Sijbrandij, M., van den Hout, M. A., & Hermans, D. (2013). Pre-trauma individual differences in extinction learning predict posttraumatic stress. *Behaviour Research and Therapy*, *51*(2), 63–67. https://doi.org/10.1016/j.brat.2012.11.004

Lonsdorf, T. B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., Jentsch, V. L., Drexler, S. M., Mertens, G., Richter, J., Sjouwerman, R., Wendt, J., & Merz, C. J. (2019). Navigating the garden of forking paths for data exclusions in fear conditioning. *ELife*, *8:e52465*. https://doi.org/10.7554/eLife.52465

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Drexler, S. M., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., … Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2017.02.026

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 97–106. https://doi.org/10.1037/0278-7393.29.1.97

Lovibond, P. F. (2011). Learning and anxiety: A cognitive perspective. In T. R. Schachtman & S. Reilly (Eds.), *Associative Learning and Conditioning: Human and non-human Applications* (pp. 104–120). Oxford University Press. http://ebooks.cambridge.org/ref/id/CBO9781107415324A009

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(1), 3–26. https://doi.org/10.1037//0097-7403.28.1.3

Luck, C. C., & Lipp, O. V. (2015). A potential pathway to the relapse of fear? Conditioned negative stimulus evaluation (but not physiological responding) resists instructed extinction. *Behaviour Research and Therapy*, *66*, 18–31. https://doi.org/10.1016/j.brat.2015.01.001

Luck, C. C., & Lipp, O. V. (2016a). The influence of contingency reversal instructions on electrodermal responding and conditional stimulus valence evaluations during differential fear conditioning. *Learning and Motivation*, *54*, 1–11. https://doi.org/10.1016/j.lmot.2016.05.001

Luck, C. C., & Lipp, O. V. (2016b). Instructed extinction in human fear conditioning: History, recent developments, and future directions. *Australian Journal of Psychology*, *68*(3), 209–227. https://doi.org/10.1111/ajpy.12135

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276–298. https://doi.org/10.1037/h0076778

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. https://doi.org/10.1016/j.biopsycho.2018.07.002

Mertens, G., & De Houwer, J. (2016). Potentiation of the startle reflex is in line with contingency reversal instructions rather than the conditioning history. *Biological Psychology*, *113*, 91–99. https://doi.org/10.1016/j.biopsycho.2015.11.014

Mertens, G., & Engelhard, I. M. (2020). A systematic review and meta-analysis of the evidence

for unaware fear conditioning. *Neuroscience & Biobehavioral Reviews*, *108*, 254–268. https://doi.org/10.1016/j.neubiorev.2019.11.012

Mertens, G., Kuhn, M., Raes, A. K., Kalisch, R., De Houwer, J., & Lonsdorf, T. B. (2016). Fear expression and return of fear following threat instruction with or without direct contingency experience. *Cognition and Emotion*, *30*(5), 968–984. https://doi.org/10.1080/02699931.2015.1038219

Mertens, G., Wagensveld, P., & Engelhard, I. M. (2019). Cue conditioning using a virtual spider discriminates between high and low spider fearful individuals. *Computers in Human Behavior*, *91*, 192–200. https://doi.org/10.1016/j.chb.2018.10.006

Milad, M. R., Orr, S. P., Pitman, R. K., & Rauch, S. L. (2005). Context modulation of memory for fear extinction in humans. *Psychophysiology*, *42*(4), 456–464. https://doi.org/10.1111/j.1469-8986.2005.00302.x

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. *The American Psychologist*, *61*(1), 10–26. https://doi.org/10.1037/0003-066X.61.1.10

Miskovic, V., & Keil, A. (2013). Visuocortical changes during delay and trace aversive conditioning: Evidence from steady-state visual evoked potentials. *Emotion*, *13*(3), 554–561. https://doi.org/10.1037/a0031323

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *The Behavioral and Brain Sciences*, *32*(2), 183–198; discussion 198-246. https://doi.org/10.1017/S0140525X09000855

Morriss, J., Christakou, A., & van Reekum, C. M. (2016). Nothing is safe: Intolerance of uncertainty is associated with compromised fear extinction learning. *Biological Psychology*,

121, 187–193. https://doi.org/10.1016/j.biopsycho.2016.05.001

Morriss, J., Saldarini, F., Chapman, C., Pollard, M., & van Reekum, C. M. (2019). Out with the old and in with the new: The role of intolerance of uncertainty in reversal of threat and safety. *Journal of Experimental Psychopathology*, *10*(1), 204380871983445. https://doi.org/10.1177/2043808719834451

Nitta, Y., Takahashi, T., Haitani, T., Sugimori, E., & Kumano, H. (2020). Avoidance Behavior Prevents Modification of Fear Memory During Reconsolidation. *Psychological Reports*, *123*(2), 224–238. https://doi.org/10.1177/0033294118811116

Olsson, A., & Phelps, E. a. (2007). Social learning of fear. *Nature Neuroscience*, *10*(9), 1095–1102. https://doi.org/10.1038/nn1968

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Pineles, S. L., Orr, M. R., & Orr, S. P. (2009). An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. *Psychophysiology*, *46*(5), 984–995. https://doi.org/10.1111/j.1469-8986.2009.00852.x

Raes, A. K., De Houwer, J., De Schryver, M., Brass, M., & Kalisch, R. (2014). Do CS-US Pairings Actually Matter? A Within-Subject Comparison of Instructed Fear Conditioning with and without Actual CS-US Pairings. *PloS One*, *9*(1), e84888. https://doi.org/10.1371/journal.pone.0084888

Raes, A. K., De Houwer, J., Verschuere, B., & De Raedt, R. (2011). Return of fear after retrospective inferences about the absence of an unconditioned stimulus during extinction. *Behaviour Research and Therapy*, *49*(3), 212–218.

https://doi.org/10.1016/j.brat.2010.12.004

Raes, A. K., De Raedt, R., Fias, W., Koster, E. H. W., & Van Damme, S. (2009). Does contingency awareness mediate the influence of emotional learning on the cueing of visual attention? *Psychological Research*, *73*(1), 107–113. https://doi.org/10.1007/s00426-008-0141-y

Reddan, M. C., Wager, T. D., & Schiller, D. (2018). Attenuating Neural Threat Expression with Imagination. *Neuron*, *100*(4), 994-1005.e4. https://doi.org/10.1016/j.neuron.2018.10.047

Rescorla, R. a. (1988). Pavlovian conditioning. It's not what you think it is. *The American Psychologist*, *43*(3), 151–160. http://www.ncbi.nlm.nih.gov/pubmed/3364852

Rowles, M. E., Lipp, O. V, & Mallan, K. M. (2012). On the resistance to extinction of fear conditioned to angry faces. *Psychophysiology*, *49*(3), 375–380. https://doi.org/10.1111/j.1469-8986.2011.01308.x

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., … Nosek, B. A. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Singh, K., Dawson, M. E., Schell, A. M., Courtney, C. G., & Payne, A. F. H. (2013). Can human autonomic classical conditioning occur without contingency awareness? The critical importance of the trial sequence. *Biological Psychology*, *93*(1), 197–205. https://doi.org/10.1016/j.biopsycho.2013.02.007

Sjouwerman, R., Niehaus, J., Kuhn, M., & Lonsdorf, T. B. (2016). Don't startle me-Interference

of startle probe presentations and intermittent ratings with fear acquisition. *Psychophysiology*, *53*(12), 1889–1899. https://doi.org/10.1111/psyp.12761

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

Tabbert, K., Stark, R., Kirsch, P., & Vaitl, D. (2006). Dissociation of neural responses and skin conductance reactions during fear conditioning with and without awareness of stimulus contingencies. *NeuroImage*, *32*(2), 761–770. https://doi.org/10.1016/j.neuroimage.2006.03.038

van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (2000). *Handleiding bij de Zelfbeoordelings Vragenlijst*. *Een Nederlandstalige bewerking van de Spielberger State-Trait Anxiety Inventory*.

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear Extinction and Relapse: State of the Art. *Annual Review of Clinical Psychology*, *9*(1), 215–248. https://doi.org/10.1146/annurev-clinpsy-050212-185542

Wegerer, M., Blechert, J., Kerschbaum, H., & Wilhelm, F. H. (2013). Relationship between fear conditionability and aversive memories: Evidence from a novel conditioned-intrusion paradigm. *PLoS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0079025

Wilson, G. D. (1968). Reversal of differential GSR conditioning by instructions. *Journal of Experimental Psychology*, *76*(3, Pt.1), 491–493. https://doi.org/10.1037/h0025540

Author Notes

GM, AMK, and IME are supported by a NWO VICI grant (grant number: 453-15-005) awarded to IME. YB is supported by Ghent University grant BOF16/MET_V/002 awarded to Jan De Houwer. We would like to thanks Inger Ebeling, Merijn van Logtestijn, Twinkle Correa, and Alexandra Klosson for their help with collecting the data. The authors declare no conflicts of interest with respect to the authorship or the publication of this article. Any correspondence regarding this article should be send to GM (g.mertens@tilburguniversity.edu).

Material and data availability

The materials, data files and data analysis scripts related to this experiment are made available through the Open Science Framework (https://osf.io/7j56p/).

Tables

Table 1. Descriptive statistics of the demographic information of the participants in the three different conditions of the experiment.

	Precise contingency instructions (n = 35)	General contingency instructions (n = 33)	No contingency instructions (n = 34)	Group comparison
Mean age in years (SD)	23.14 (3.17)	23.09 (3.23)	23.53 (3.57)	$F(2, 99) < 1$
Gender distribution	25 females 10 males	21 females 12 males	24 females 10 males	$X^2(2) < 1$
Mean STAI-T (SD)	42.23 (8.66)	41.55 (8.96)	40.18 (10.48)	$F(2, 99) < 1$
Mean US intensity in mA (SD)	4.11 (3.64)	4.20 (2.60)	5.46 (6.48)	$F(2, 99) < 1$
Mean US rated pain on a 0-10 scale (SD)	5.50 (0.92)	5.43 (0.96)	5.75 (1.07)	$F(2, 99) < 1$

Table 2. Fear acquisition and contingency awareness rates for the different conditions in the experiment.

	Precise contingency instructions (n = 35)	General contingency instructions (n = 33)	No contingency instructions (n = 34)	Group comparison
% successful SCR discrimination	68.6% (24 out of 35)	63.6% (21 out of 33)	52.9% (18 out of 34)	$\chi^2(2) = 1.86$ (ns)
% successful FPS discrimination	60.0% (21 out of 35)	81.8% (27 out of 33)	67.65% (23 out of 34)	$\chi^2(2) = 3.92$ (ns)
% contingency aware acquisition phase	71.4% (25 out of 35)	90.9% (30 out of 33)	52.9% (18 out of 34)	$\chi^2(2) = 11.87$*
% contingency aware reversal phase	48.6% (17 out of 25)	60.6% (20 out of 33)	47.1% (16 out of 34)	$\chi^2(2) = 1.48$ (ns)

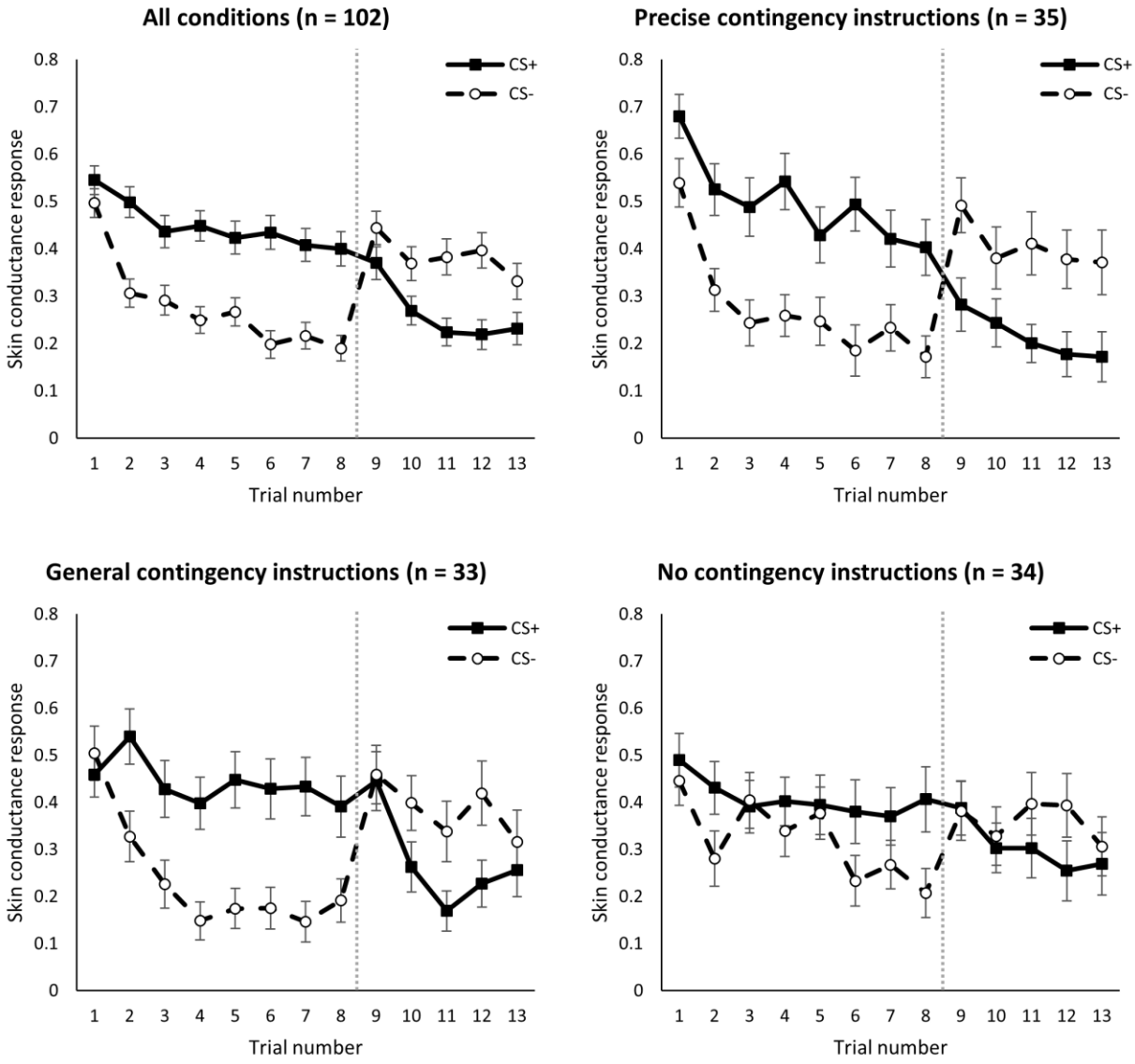
*p = .003

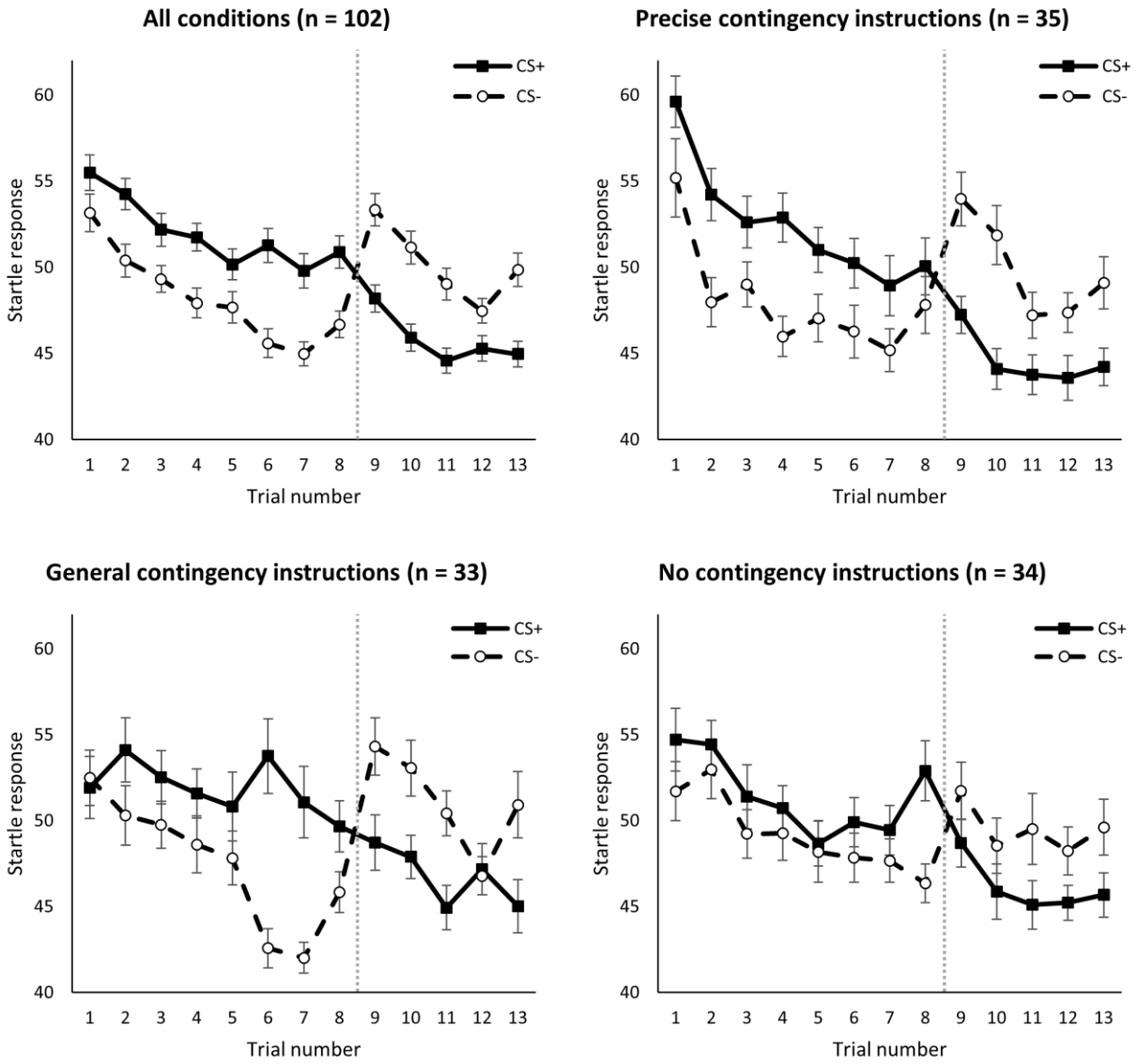
Figure Captions

Figure 1. Range corrected and square root transformed skin conductance responses with the whole sample (top left panel) and across the three different conditions (top right and bottom panels). The dashed lines indicate when contingency reversal instructions were given. Error bars display the standard error of the mean.

Figure 2. T-transformed startle responses throughout the experiment with the whole sample (top left panel) and across the three different conditions (top right and bottom panels). The dashed lines indicate when the contingency reversal instructions were given. Error bars display the standard error of the mean.

Figures


All conditions (n = 102)
Precise contingency instructions (n = 35)
General contingency instructions (n = 33)
No contingency instructions (n = 34)
CS+
CS-
Skin conductance response
Trial number
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1 2 3 4 5 6 7 8 9 10 11 12 13


All conditions (n = 102)
Precise contingency instructions (n = 35)
General contingency instructions (n = 33)
No contingency instructions (n = 34)
CS+
CS-
Startle response
Trial number
60
55
50
45
40
1 2 3 4 5 6 7 8 9 10 11 12 13