**An attentional perspective on differential fear conditioning in chronic pain: The
informational value of safety cues.**

Stephan Koenig^{1,2}, Karoline Körfer¹, Harald Lachnit¹ & Julia Anna Glombiewski²

Author Note ¹ Department of Psychology, Philipps-Universität Marburg.

² Department of Psychology, Universität Koblenz-Landau.

Correspondence concerning this article should be addressed to Stephan Koenig
(koenigst@uni-landau.de)

Abstract

Differences in fear conditioning between individuals suffering from chronic pain and healthy controls may indicate a learning bias that contributes to the acquisition and persistence of chronic pain. However, evidence from lab-controlled conditioning studies is sparse and previous experiments have produced inconsistent findings. Twenty-five participants suffering from chronic back pain and twenty-five controls not reporting chronic pain took part in a differential fear conditioning experiment measuring attention (eye tracking) and autonomic arousal (pupil dilation and skin conductance) elicited by visual cues predicting the presence or absence of electric shock. In contrast to the healthy control group, participants with chronic pain did not acquire differential autonomic responding to cues of threat and safety and specifically failed to acquire any attentional preference for the safety cue over irrelevant contextual cues (while such preference was intact for the threat cue). We present simulations of a reinforcement learning model to show how the pattern of data can be explained by assuming that participants with chronic pain might have experienced less positive emotion (relief) when the electric shock was absent following safety cues. Our model shows how this assumption can explain both, reduced differential responding to cues of threat and safety as well as less selective attention to the safety cue.

Keywords: Chronic Pain, Fear Conditioning, Impaired safety learning, Informational value, Attention, Eye Movements.

An attentional perspective on differential fear conditioning in chronic pain: The informational value of safety cues.

Associative learning has been acknowledged to play a central role in the acquisition and preservation of chronic pain (Linton et al., 1984; Madden & Moseley, 2016; Meulders, 2019, 2020; Moseley & Vlaeyen, 2015; Vlaeyen, 2015; Vlaen & Linton, 2012; Zaman et al., 2015) as well as comorbid anxiodepressive disorders (Duits et al., 2015; Lissek et al., 2005; Yalcin & Barrot, 2014). In a prototypical learning episode a person might perform a specific type of bodily movement that is followed by an acute episode of pain. In terms of classical conditioning, the proprioceptive feedback from the movement takes the functional role of a conditioned stimulus (CS) that is followed by pain as an aversive unconditioned stimulus (US). The pairing establishes a CS-US-association in memory which causes movement-related fear and avoidance of such movement in the future. Finally, generalized fear spreads to other movements and leads to dysfunctional and widespread avoidance behavior and disability as major symptoms in chronic pain.

If the transition from acute to chronic pain indeed is based on associative processes as exemplified above, individuals with chronic pain might exhibit a specific learning bias that facilitates the acquisition and generalization of pain-related fear. *Differential fear conditioning* has been used as an empirical model to study these processes under controlled conditions in the laboratory. Here, participants repeatedly observe that some initially neutral CS (e. g. pictures, tones, movements) are associated with a pain US (e. g. electric shock) and thus constitute a threat cue (CS+) while other CS are explicitly paired with the absence of pain and thus constitute a safety cue (CS-). Learning this discrimination typically results in stronger conditioned

54 responding (as measured by skin conductance, blink startle, heart rate, pupil dilation etc.) to the
CS+ than the CS- indicating that the former elicits more conditioned fear than the latter.

56 In a systematic review of studies that employed such procedures to examine differential
conditioning in chronic pain, Harvie et al. (2017) concluded that there was some preliminary
58 evidence for a learning deficit in individuals with chronic pain, who exhibited smaller
differential responding than healthy controls on some measures. The authors suggested that such
60 deficit may be based on a failure to establish the CS- as a cue that signals safety from harm in an
otherwise threatening context. In accord with this hypothesis of *impaired safety learning* some
62 studies have reported elevated responding to the CS- in individuals with chronic pain during fear
conditioning (Jenewein et al. 2013; Meulders et al. 2014; but see Icenhour et al., 2015; Klinger et
64 al., 2010) as well as marked generalization of fear to stimuli similar to the CS- after conditioning
(Meulders et al., 2015; Meulders et al., 2017).

66 We conducted the current experiment to confirm the hypothesis of *impaired safety learning*
in chronic pain and specifically to examine *attentional correlates* of (failed) differential fear
68 conditioning. We hypothesized that measures of selective attention (to the CS+, the CS-, and the
conditioning context) would give further information about two possible mechanisms underlying
70 the over-generalization of fear in chronic pain: On the one hand, while the CS+ typically draws
more attention than the CS- in healthy participants (Koenig, Uengoer et al., 2017; Notebaert et
72 al., 2011; Van Damme et al., 2004; Wang et al. 2013), attention to the CS- might be elevated in
chronic pain if there is strong generalization of fear from the CS+ and the CS- is thus treated as a
74 presumed threat cue (Meulders et al., 2015; Meulders et al., 2017). Such a pattern could be
caused by stronger sensitization processes in participants with chronic pain that cause stronger
76 responding to *all* stimuli presented in the conditioning context. On the other hand, over-

generalization might be accompanied by decreased attention to the CS if chronic pain patients fail to register that these CS provide the relevant perceptual cues to predict the US and instead focus on irrelevant contextual stimuli. With respect to chronic pain, when a patient experiences that some movements cause acute pain while others do not, these real-life learning episodes typically also include a wide array of further interoceptive and exteroceptive contextual cues. In such a cluttered perceptual environment any failure to identify and attend to discriminative, unique features of the CS might provide the basis for retarded discrimination learning causing overgeneralization and free-floating, contextual fear of movement.

The fundamental role of attention in classical conditioning has long been acknowledged by formal theories of associative learning (Mackintosh, 1975; Pearce and Hall, 1980; Pearce and Mackintosh 2010; Le Pelley 2004). Converging empirical evidence points to the fact that observing the contingency between a CS and a US will not only results in a learned CS-US association to drive expectancy and conditioned responding but will also result in a learned attentional bias for the CS that affects attention for liking, action and learning in the future (Gottlieb, 2012; Le Pelley et al. 2016).

To examine attentional learning in participants with chronic pain and healthy controls we used a fear conditioning paradigm in which localized visual CS were predictive of the occurrence or omission of a subsequent electric shock but these CS were always embedded in a search display of irrelevant contextual stimuli (that were not correlated with shock but competed for attentional selection). We measured skin conductance and pupil dilation to index conditioned emotional arousal elicited by the search display and used eye tracking to examine the distribution of visual attention across the relevant CS (CS+ and CS-) and the irrelevant contextual stimuli. In a subsequent test stage participants had to search for a newly designated target stimulus while

shocks were no longer administered. The CS from the conditioning stage were introduced as task-irrelevant distractors (that had to be ignored for best performance) and we examined the distractors' potential to capture and hold attention automatically (Anderson, 2015; Koenig et al., 2017; Wentura et al., 2014; Schmidt et al., 2015a, 2015b). If attentional learning during fear conditioning appears to be altered in participants with chronic pain, this finding might provide an important new perspective for future models of chronic pain as well as therapeutic treatment.

Method

The study was approved by the local ethics committee of the Faculty of Psychology at the Philipps-Universität Marburg (AZ: 2018-12). All subjects gave written informed consent in accordance with the Declaration of Helsinki. Methods and hypotheses were preregistered at <https://aspredicted.org/65b7u.pdf>.

Participants

Participants were recruited via newspapers, mailing lists and flyers at pharmacies, supermarkets and local hospitals. They were included if their age ranged between 18 and 65 years and they did not report a lifetime psychotic disorder or a current drug abuse. Participants were included in the experimental group if they reported chronic pain (≥ 3 months) primarily in the lower back on more than 3 days a week as a major complaint. The experimental group consisted of twenty-five participants with chronic back pain. The control group included twenty-five participants not reporting pain in the last 6 months that were matched for sex and age. Sample size was based on *a priori* power calculation with G*Power 3.1.6 (Faul et al., 2007) suggesting a total sample size of $N = 48$ to detect small to medium effects ($f = .20$, $\eta_p^2 = 0.40$) of

impaired differential fear conditioning in participants with chronic pain (Harvie et al., 2017) with
124 a statistical power of $1 - \beta = .85$, type I error $\alpha = .05$, critical $F(2, 92) = 3.095$.

All participants received monetary compensation (40 EUR) and had normal or corrected-
126 to-normal vision. Demographic characteristics are displayed in Table 1. Participants in the
experimental group indicated an average pain intensity of $M = 5.14$ ($SD = 1.96$) on an 11-point
128 numeric rating scale ("Please indicate your average pain intensity in the last 4 weeks", 0 = "no
pain"; 10 = "worst imaginable pain"; Nagel, Gerbershagen, Lindena, & Pfingsten, 2002) for the
130 last 4 weeks preceding the experiment. The average pain duration was 3.24 ($SD = 22.34$) years
and the average disability, assessed with the Pain Disability Index (Pollard, 1984), was 28.38
132 ($SD = 12.69$).

134 Apparatus

Eye movements were acquired using an infrared eye tracker (Eyelink 2000, SR-Research,
136 Canada) which sampled gaze position and pupil size at 1000Hz. The apparatus restrained the
participants head via chin and forehead rests and was table-mounted in front of a 22"-CRT
138 monitor (Iiyama, Vision Master Pro514) with an eye-to-screen-distance of 78 cm. The eye
tracker was calibrated to yield a maximal error $< 0.5^\circ$. Palmar skin conductance was recorded at
140 50 Hz using a constant voltage bridge (Lykken & Venables, 1971) interfacing custom
LabVIEW® software (National Instruments, Austin, TX, USA). 8 mm silver-silver chloride
142 electrodes were filled with an electrolytic medium (Unibase, with 0.05 mol NaCl) and were
attached to the thenar and hypothenar eminences of the participant's left hand. The experiment
144 was controlled by Presentation® software (Neurobehavioral Systems, USA). A 10 ms dc electric
shock was used as the aversive unconditioned stimulus and was delivered via silver-silver

chloride electrodes to the volar surface of the participant's right arm from an isolated transformer–condensor shock generator (Kimmel et al., 1980).

Stimuli

Two sets of stimulus colors were chosen to be equidistant in CIE L^*a^*b color space and were matched for lightness ($L^* = 60$) and chroma ($C^* = 34$). Hues in the first set were $h^\circ = 30$ (red), 150 (green), and 270 (blue) respectively. Hues in the second set were $h^\circ = 90$ (yellow), 150 (cyan), and 270 (magenta). Gray distractor stimuli and a fixation cross at the start of each trial were matched for lightness ($L^* = 60$, $C^* = 0$). All stimuli were presented on a light gray ($L^* = 87$) background. As depicted in Figure 1a, during fear conditioning a colored annulus and a colored diamond were presented among four gray distractors for 5 seconds sometimes followed by the aversive electric stimulus. All stimuli were evenly distanced from screen center (100 mm). In each trial the colored annulus was rendered in one of three colors (e. g. red, green or blue) and was the relevant cue for predicting shock. The diamond was rendered in one of three different colors (e. g. yellow, cyan, magenta) and provided a contextual cue that was irrelevant for predicting shock. For one half of the participants the relevant annulus colors were randomly drawn from the first color set and irrelevant diamond colors were drawn from the second set while the reverse was true for the other half of the participants.

Figure 1b depicts the stimuli presented in search trials. The designated target stimulus was a gray diamond containing a small line that was oriented either vertically (to instruct participants to respond with a left-button mouse click) or horizontally (for a right-button click). Oblique lines were presented inside the distractor annuli but were irrelevant in the search task. In 75% of the

168 trials one distractor annulus was rendered in one of the relevant cue colors from the conditioning stage.

170 -----
Insert Figure 1 about here
172 -----

174 Procedure

Prior to the experiment written instructions were presented and participants gave informed consent. The experimenter attached recording and stimulation electrodes and delivered shocks of increasing intensity for participants to choose an intensity that was “definitely unpleasant but not really painful”. The adjustment of stimulation strength to an individual tolerance threshold did not differ between groups of patients and controls, $F < 1$. The eye tracker was calibrated and the experiment started with 45 trials of aversive conditioning that were followed by 240 search trials.

Aversive conditioning. Each trial started with a 2-sec fixation cross that instructed participants to stop blinking and pay attention (Figure 1a). In the following 5-sec search display participants were presented with a colored annulus and a colored diamond to predict whether an aversive electric shock would be administered when the search display disappeared. The annulus color was the *relevant* cue for predicting shock and across trials three different colors were conditioned to elicit high (H), partial (P), or low (L) shock expectancy. The diamond was rendered in one of three additional colors X, Y, and Z to provide a contextual cue that would draw attention because of its physical salience but that was *irrelevant* for predicting shock.

Specifically, combinations HX, HY, and HZ were always followed by shock, combinations PX, PY, and PZ were sometimes followed by shock (on 50% of the trials), and combinations LX,

LY, and LZ were never followed by shock. Within each of three consecutive blocks of
192 conditioning trials, each relevant color was presented six times resulting in 54 conditioning trials
in total, with 18 presentations of each relevant color. Trials were presented in pseudo-random
194 sequence with the restriction that trials with an electric shock would not exceed three
presentations in a row.

196 *Visual search task.* After conditioning, participants performed a search task consisting of
240 trials (Figure 1b). Each trial started with a 2-sec fixation cross. In the subsequent search
198 display participants had to find and fixate on the gray diamond as quickly as possible to identify
the embedded line orientation by pressing a left (vertical line) vs. right (horizontal line) button.
200 Line orientation was a high acuity feature that could only be identified if participants moved
their eyes to fixate on the diamond target. On 75% of the search trials, the display included a
202 color from the conditioning stage (H, P, L) as a distractor to interfere with the task. An initial
fixation on the color distractor delayed the time participants arrived at the target and were able to
204 identify the embedded line orientation. At the end of each search trial the reaction time was
presented within the diamond target to motivate quick responding. If participants pressed the
206 wrong button a capital “F” (for “False”) was shown within the diamond. Search trials were
administered in five consecutive blocks of 48 trials in random sequence. Participants were
208 instructed that there would be no further shocks in the search task, that they had to respond to the
shape target as quickly as possible, and that “all circles in the task could be ignored”.

210 **Dependent variables**

212 We used custom MATLAB software (The MathWorks, Inc., 2012; Koenig 2010) for the
processing of eye position and pupil size and the velocity-based detection of ocular fixations

(threshold 30°/sec). Blink artifacts in pupil traces were corrected using cubic spline interpolation. Signal averaging and statistical analysis was conducted using the R language and environment for statistical computing (R Core Team, 2020).

Pupil size. In conditioning trials, we computed the average dilation of the pupil from 0 to 5000 ms after CS onset to indicate autonomic arousal in anticipation of shock (Lonsdorf et al., 2017). In each trial the pupil size at CS onset was subtracted from all samples in the pupil trace to center the trace at the event of interest. The resulting trace was then divided by initial pupil size at CS onset (and multiplied by 100) to yield a measure of percentual change (see Koenig, Uengoer et al., 2017, 2018; Preuschhoff et al., 2011). Traces for each trial were first averaged within subject within experimental condition and secondly across subjects within condition to yield average pupil traces.

Skin conductance response. Skin conductance changes in micro Siemens (mS) were transformed to standardized z-values and averaged across trials within condition as for pupil size. Please note that the skin conductance response was not mentioned in the preregistration of the experiment. We included the skin conductance response subsequently because it was readily available in the lab as a second measure of autonomic arousal. We interpret both, pupil dilation and skin conductance to reflect activity in the sympathetic branch of the autonomic nervous system and thus to indicate the activation of the fight-or-flight system in response to shock-associated CS (threat cues) in the context of our fear conditioning experiment. We do not have any psychological hypothesis about why the two measures should differ and model simulations of conditioned fear relate to both variables in equal measure.

Fixation probability and total dwell time. In conditioning trials, participants moved their eyes to explore the search display and to inspect the colored annulus (relevant color), the

diamond shape (irrelevant color), or one of the circular gray distracters. To provide a proxy of
overt attention, we computed the continuous probability of fixating these different stimuli over
the 5-sec CS interval. For each millisecond in the eye position trace we scored a value of 1 if
gaze was directed to the CS (within 30° of the CS position) and a value of 0 otherwise.

Averaging these scores across trials yielded continuous fixation probability at every millisecond
following CS onset. The area under this probability trace in an interval from 0 to 5 sec following
CS onset indicates total dwell time in the CS interval (and is equivalent to summing durations of
single fixations). Fixation probability and total dwell time for the irrelevant contextual cue was
computed accordingly.

Frequency of attentional capture. In search trials, participants were instructed to find and
fixate on the shape target and to ignore any simultaneous color distractor. We computed the
relative frequency of search trials in which the participants first fixation nonetheless was on the
color distractor. Capture frequency has been shown to correspond to distractor value acquired by
association with an appetitive or aversive outcome (Koenig, Kadel et al., 2017; Koenig, Uengoer
et al., 2017).

Duration of attentional capture. If the CS color distractor automatically captured the first
fixation in search trials, we computed the duration of that fixation (until participants disengaged
from the distractor to move their eyes to the target).

CS memory and contingency awareness. At the end of the experiment participants indicated
on a questionnaire if they believed that there had been a systematic relationship between colors
in the learning task and electric shock (translated from German: “In this experiment participants
have been randomly assigned to different groups. In some groups there was a systematic
relationship between the circle colors and the electric stimulation. Do you believe that you were

in a group with such a systematic relationship?”). They were asked to recall and write down the colors from the conditioning stage and describe the colors’ association with shock in an open response format (translated from German: “If you recognized a correlation between the colors and the electric stimulation, please try to describe now:”). Contingency knowledge was scored as correct if a participant correctly recalled the CS color and correctly indicated whether it was (sometimes) followed by electric stimulation (CS H and P) or never followed by electric stimulation (CS L). Please note that with our primary interest in attentional measures the questionnaire was given after an extensive number of extinction trials in the search task (presenting the CS colors without the US) and thus does not constitute a typical post-acquisition test.

Hypothesis and statistical analysis

Hypothesis 1: Impaired discrimination of threat and safety cues. Participants suffering from chronic pain should exhibit less differential fear conditioning than healthy controls. The difference in responding to the CS+ color combinations (HX+, HY+, HZ+) compared to the CS- color combinations (LX-, LY-, LZ-) thus should be smaller in participants with chronic pain than in the healthy control group.

Hypothesis 2: The CS- is treated as presumed threat cue and draws attention. In trials LX-, LY-, and LZ- participants with chronic pain could misidentify cue L as yet another threat cue in the conditioning context because of strong generalization from cues H and P. Since feared cues typically draw more attention than non-feared cues, this should selectively increase attention to cue L in participants with chronic pain.

Hypothesis 3: Increased attention to irrelevant cues and impaired attention to relevant cues. Overgeneralization of fear in chronic pain might be based on increased attention to common elements X, Y, Z at the cost of decreased attention to the relevant elements L, P, and H. Fixation probability and duration thus could be decreased in participants with chronic pain.

Caveat: Hypothesis 2 (more attention to the CS-) and Hypothesis 3 (less attention to the CS-) formulate two mutually exclusive hypothesis about the nature of impaired differential fear conditioning in chronic pain. Hypotheses were preregistered at <https://aspredicted.org/65b7u.pdf>.

For statistical analysis, we used apriori contrasts to test Hypothesis 1 of reduced differential fear learning in chronic pain patients. Across conditions *low*, *partial*, and *high*, coefficients [-1, 0, 1] encoded the hypothesis of more autonomic arousal (pupil dilation, skin conductance) in trials that elicited high shock expectancy than in trials that elicited low shock expectancy. Our interaction hypothesis was that this contrast was smaller in chronic pain patients than in healthy controls [-1, 1]. Tests for these directed hypotheses were one-sided. In contrast, attentional hypotheses were not derived from previous findings but were rather based on principled consideration. To explore attentional allocation during and after learning, we used a general linear model with factors Group (Pain, Control), Threat (Low, Partial, High), and Stimulus (Relevant, Irrelevant). Omnibus *F*-Tests in these exploratory analysis were two-sided. Violations of sphericity were corrected using the method of Huynh and Feldt (1976). Effect sizes are reported as partial eta squared (for ANOVAs), contrast correlation, or Cramer's *V* (for Chi-square tests).

Results

Fear conditioning

Shock intensity. Intensity of aversive electric stimulation was adjusted to an individual level that was “definitively unpleasant” to the participant prior to the experiment. Average intensities of constant direct current stimulations were 1.642 mA ($SE = .474$) in the experimental group and 1.136 mA ($SE = 0.464$) in the control group, with no difference between groups, $F < 1$.

Pupil dilation. Figure 2a depicts increases in pupil size elicited by the onset of the CS in chronic pain patients (left panel) and the matched control group (right panel) during fear conditioning. Correspondingly, Figure 2b depicts pupil size averaged across the entire 5-sec CS interval. Pupil dilation in the control group was more pronounced when the search display contained a high threat cue (H) than when it contained a partial threat cue (P) or a low threat cue (L). A general linear model coding for factors of CS (L, P, H) and Group (Pain, Control) revealed no main effect of Group, $F < 1$, but an effect of CS, $F(2,94) = 10.229$, $p < .001$, $\eta^2_p = .176$, that was modulated by a Group X CS interaction, $F(2, 94) = 3.286$, $p = .042$, $\eta^2_p = .064$ (please see supplement Figure S1 for how changes in pupil size evolved during fear conditioning). An a priori CS contrast [-1, 0, 1] coding for stronger pupil dilation elicited by the high threat cue than the low threat cue L was highly significant in the control group, $t(96) = 4.901$, $p < .001$, $r = .447$, but fell short of statistical significance in participants with chronic pain, $t(96) = 1.358$, $p = .085$, $r = .139$. A Group x CS interaction contrast [-1, 1] confirmed our a priori hypothesis that differential responding to high threat vs. low threat cues was more pronounced in the control group than in the experimental group, $t(96) = 2.486$, $p = .007$, $r = .246$. Further analysis revealed that these differences in conditioned pupil dilation were not caused by different pre-stimulus baselines: At the time of CS onset there were no significant differences in pupil size between participants in the experimental group and the control group, $F < 1$.

Insert Figure 2 about here

Skin conductance. The experiment also included the skin conductance response as a second measure of autonomic emotional arousal. Results are shown in Figure 3. A general linear model coding for factors revealed no main effects of Group, $F(1,47) = 2.648, p < .110, \eta^2_p = .053$, and CS, $F(2,82) = 1.583, p < .213, \eta^2_p = .176$, and no interaction, $F(2, 82) = 1.973, p = .151, \eta^2_p = .040$ (please see the supplement Figure S2 for how changes in skin conductance evolved during fear conditioning). However, a planned contrasts comparing low and high threat cues was significant in the control group, $t(94) = 2.363, p = .010, r = .237$, but not in the experimental group, $t(94) = -.284, p = .611, r = .029$. Importantly, the planned interaction contrast was significant, $t(40) = 1.689, p = .049, r = .258$, and confirmed that differential responding was more pronounced in the control group than in the experimental group. However, conditioned skin conductance responses were set against different levels of pre-stimulus baselines. During the last second preceding CS onset the average skin conductance level (SCL) was $4.889 \mu\text{S}$ ($SE = .241$) in the experimental group and was significantly reduced compared to an SCL of $6.710 \mu\text{S}$ ($SE = .349$) in the control group, $F(1,47) = 6.111, p = .017, \eta^2_p = .115$.

Insert Figure 3 about here

Fixation probability and dwell time. During conditioning, the search display was presented for 5 seconds in each trial and Figure 4 depicts the probabilities of fixating on the items in the search array during this interval. The shaded area in Figure 4 represents the extent to which the probability of fixating the relevant threat cue (colored annulus) exceeded the probability of fixating on the irrelevant distractor (colored diamond) and corresponds to differences in total dwell time as shown in Figure 5. An ANOVA with factors stimulus (cue, distractor), threat (low, partial, high) and Group (Pain, Control) revealed main effects of group, $F(1, 48) = 4.705, p = .035, \eta^2_p = .089$, stimulus, $F(1, 48) = 29.080, p < .001, \eta^2_p = .377$, and threat, $F(2, 87) = 4.142, p = .022, \eta^2_p = .079$. Interactions were Group X Threat, $F(2, 87) = 3.034, p = .058, \eta^2_p = .059$, Stimulus X Threat, $F(2, 84) = 15.892, p < .001, \eta^2_p = 0.248$, and Stimulus X Threat X Group, $F(2, 84) = 3.270, p = .049, \eta^2_p = .063$. The three-way interaction was caused by the fact that in chronic pain patients the attentional preference of the cue over the distractor was more strongly modulated by threat level than in healthy controls. Specifically, dwell-time differences (cue minus distractor) were $M_L = 270\text{ms}, |t| < 1, M_P = 906\text{ms}, t(62) = 2.939, p = 0.004$, and $M_H = 1348\text{ms}, t(62) = 4.370, p < .001$, for levels of low, partial and high threat respectively (pooled $SE = 308\text{ms}$ for all pairwise comparisons) in chronic patients, $F(1, 35) = 11.924, p < .001, \eta^2_p = .332$. In contrast, the same differences were 1147ms, 1387ms, and 1552ms (all $p < .001$), $F(2, 46) = 4.008, p = .026, \eta^2_p = .143$ in the healthy control group. Attentional preference for the high threat cue over the irrelevant distractor thus was intact for both groups, but in contrast to the control group the pain group specifically failed to exhibit such preference for the safety signal (low threat cue).

Insert Figures 4 and 5 about here

Search task

Capture frequency. In each search trial, participants had to find and fixate the gray diamond target as quickly as possible while colored circles (colors L, P, and H from the conditioning stage) were shown as task-irrelevant distractors that had to be ignored in order to perform well. Figure 6 depicts the frequency of trials in which our participants first fixation after onset of the search display was on the color distractor (before eventually looking at the diamond shape target). The figure shows that capture frequency was dependent on the value that the distractors had acquired during fear conditioning. Participants were most likely to fixate on the distractor if it was a high threat cue during fear conditioning while such capture was less likely for partial and low threat cues. An ANOVA revealed a main effect for distractor type, $F(2,93) = 7.769, p < .001, \eta^2_p = .139$, but no effect for group and no interaction (both $F < 1$). Linear contrasts [-1, 0, 1] for the increase in capture frequency with shock expectancy (low, partial, high) were significant for both, the experimental group, $t(96) = 3.035, p = 0.001, r = .296$, and the control group, $t(96) = 2.537, p = 0.006, r = .251$.

Capture duration. In capture trials, dwell time on the color distractor (before eventually looking at the target) was 152ms ($SE=4.38$) on average but was affected by conditioned levels of threat as well. In the experimental group the duration of capture by colors L, P, and H was 146ms, 151ms, and 155ms respectively. In the control group these were 151ms, 153ms, and

159ms. The main effect of CS was $F(2,59) = 2.941, p = .061, \eta^2_p = .130$, with no effect of group
and no interaction (both $F < 1$).

Insert Figure 6 about here

Post experimental CS memory and contingency awareness.

At the end of the experiment participants indicated on a questionnaire if they believed that
there had been a systematic relationship between colors in the learning task and electric shock.

After the experiment 76% of participants with chronic pain and 72% of participants in the
control group believed in such a relationship with no difference between groups, $\chi^2(1) = 0.027, p$
 $= .869, V = .023$. However, a group difference did exist with respect to *actual* CS memory
contingency knowledge. In the pain-free control group 40%, 68%, and 72% of the participants
were able to correctly remember the colors of low, partial and high threat CS from the learning
task. In contrast these percentages were 28%, 44%, and 52% for participants with chronic pain.

The main effect of group, $\chi^2(6) = 6.578, p = .010, V = .387$, indicated decreased contingency
awareness in participants with chronic pain. A main effect of CS was caused by the fact that
more participants correctly remembered the color from the high threat condition than the color
from the low threat condition, $\chi^2(5) = 27.061, p < .001, V = .775$. The Group x CS interaction
was $\chi^2(5) = 4.894, p = .086, V = .329$.

418 **Model simulations**

Figure 7 presents simulations of a reinforcement learning model further detailed in the supplement to this article. The model uses the standard learning rule of Rescorla and Wagner (1972; Wagner & Rescorla, 1972; RW model)

$$422 \quad \Delta V_i = a_i \beta (\lambda - \Sigma V) \quad (1)$$

were the change in associative strength V_i (between CS i and the aversive US) is driven by the prediction error $\lambda - \Sigma V$ (difference of observed US strength λ minus US strength predicted by all CS present ΣV) weighted by learning rates β (referring to US properties) and a_i (referring to attention to CS i). The notion of attentional learning extends the standard RW model by assuming that attention to CS i changes during conditioning

$$428 \quad a_i = \alpha_i + \omega_1 |V_i| + \omega_2 \varepsilon_i \quad (2)$$

where α_i corresponds to bottom-up salience, $|V_i|$ corresponds to the *predictiveness* of the CS for the presence or absence of the US (see Mackintosh, 1975) and changes according to Equation 1, and ε_i corresponds to the *uncertainty* of the CS in predicting the US (see Pearce & Hall, 1980) and changes according to

$$\Delta \varepsilon_i = \gamma (|\lambda - \Sigma V| - \varepsilon_i) \quad (3)$$

where ε_i is continuously driven towards the absolute value of the prediction error $|\lambda - \Sigma V|$ observed after CS i and γ is an attentional learning rate.

Weights ω_1 and ω_2 in Equation 2 are free parameters to code for the relative contribution of attention to predictive versus uncertain CS. Lastly, we assume that measures of overt attention in our experiment (fixation probability, total dwell time) correspond to a standard softmax normalization of learned attentional weights a_i

$$440 \quad f(a_i) = \exp(\varphi a_i) / \Sigma \exp(\varphi a) \quad (4)$$

Insert Figure 7 about here

Figure 7 shows simulations for a parameter set of $\alpha = .3$ for all cues, $\lambda = 1$ and $\beta = .3$ for excitatory learning (US present), $\beta = .1$ for inhibitory learning (US absent), $\gamma = .1$ (for all cues) and relative weights $\omega_1 = .3$ and $\omega_2 = .1$. The bottom panels depict predicted fear associations for cues L, P, and H in the experimental group (left) and the control group (right). More differential fear (H minus L) in the control group results from the assumption that the omission of electric shock generates a stronger teaching signal for inhibitory learning in the control group ($\lambda_{\text{noUS}} = -.3$) than in the experimental group ($\lambda_{\text{noUS}} = 0$) and matches the pattern of results observed for conditioned fear as indexed by pupil size (Figure 2) and skin conductance (Figure 3). The top panels depict selective attention predicted by the same simulation. The model exhibits less attention to the safety cue in the experimental group because the absolute value of the inhibitory CS association $|V_L|$ matches the excitatory association of the irrelevant context $|V_{\text{context}}|$. In contrast the stronger inhibitory teaching signal in the control group creates absolute weights $|V_L| > |V_{\text{context}}|$ which affect attention accordingly (Equation 2) matching the pattern of results observed for fixation probability (Figure 4) and total dwell time (Figure 5). The supplement gives further detail on the proposed mechanism.

Discussion

In accord with previous findings our results revealed that individuals suffering from chronic pain exhibited impaired differential fear learning (Duits et al., 2015; Meulders et al., 2014, 2015; Schneider et al., 2004). Levels of sympathetic autonomic arousal indexed by pupil dilation and skin conductance were higher for the CS+ (always followed by electric shock) than for the CS- (never followed by shock) in the pain-free control group, but such a differential fight-or-flight response was absent in individuals suffering from chronic pain. If we take these measures to indicate fear memory these findings suggest that participants with chronic pain failed to acquire the discrimination between cues of threat and safety which corresponds to the prolonged and free-floating threat anticipation reported for chronic pain by previous studies (Baas et al., 2008; Grillon et al., 2002; Vansteenwegen et al., 2008). As outlined in the introduction, a prominent hypothesis to explain such failed discrimination in the anxiety disorders (Duits et al., 2015; Lissek, et al. 2005) and chronic pain (Harvie et al., 2017) is based on *impaired safety learning* and assumes that patients may exhibit a failure to establish the CS- as a cue that signals safety from harm in an otherwise threatening context. However, our measures of autonomic arousal did not seem to indicate that failed discrimination was specifically based on an increased fight-or-flight response to the CS-.

Instead, our results revealed evidence for an essential *attentional contribution to impaired safety learning in chronic pain*: The major focus of our experiment was to examine the allocation of overt attention when cues of threat and safety that were *relevant* for predicting shock (cues H, P, and L eliciting high, partial, and low shock expectancy, respectively) were embedded with contextual cues (X, Y, and Z) that were *irrelevant* for the prediction. Our results revealed, that fixation probability and total dwell-time were actually *more* strongly modulated by the CS-US

contingencies in participants with chronic pain than in pain-free controls: In both groups, 488 attentional preference for the threat cue H over the irrelevant cues X, Y, and Z was higher than the preference for the safety cue L over the irrelevant cues but this difference was more 490 pronounced in participants with chronic pain (see Figures 4 and 5). This was due to the fact that participants with chronic pain did allocate more attention to the threat cue than the irrelevant 492 contextual cue but specifically failed to do the same for the safety signal: When cue L was presented in trials LX-, LY-, and LZ- the fixation dwell-time of participants with chronic pain 494 was equally distributed between the safety signal (L) and the irrelevant contextual cues (X, Y, or Z). In contrast, attentional preference for the safety cue over the irrelevant contextual cues was 496 intact in participants in the healthy control group. These findings suggest an essential attentional contribution to impaired safety learning (Harvie, et al., 2017; Meulders, 2019, 2020) and to our 498 knowledge have not been reported previously.

To further explore possible mechanisms underlying conditioned fear and attention 500 our results included simulations of a reinforcement learning model that exhibits the same behavior as our participants: Differences in attention to the CS+ versus CS- (corresponding to 502 fixation probability and total dwell time) were *more* pronounced in the group with chronic pain (Figure 7, top panels) while differences in conditioned fear elicited by the CS+ versus CS- 504 (corresponding to skin conductance and pupil dilation) were *less* pronounced with chronic pain (Figure 7, bottom panels). This pattern emerged from the single modeling assumption that the 506 omissions of the US after the CS- (the absence of electric shock) generated a teaching signal for inhibitory learning that was less effective in participants with chronic pain. Numerically, 508 simulations used a teaching signal of $\lambda = 1$ to represent the effective strength of shock after the CS+ in both groups. However, the absence of shock was modeled with $\lambda = 0$ in the chronic pain

group but $\lambda = -.3$ in the healthy control group. Psychologically, this difference may relate to the possibility that pain-free controls were more adept to derive positive emotion (e. g. relief) from the absence of the aversive US. This notion in principle is consistent with empirical studies showing that conditioned safety signals may acquire independent rewarding properties (Christianson et al., 2012; Ganguly & Kleinfeld, 2004; Green et al, 2018; Pollak et al., 2008; Rogan et al., 2005). It would mean that in healthy controls the summed prediction from perceiving the CS- in the aversive conditioning context elicited positive emotion to some extent with no such positive emotion in chronic pain patients.

With respect to our interpretation of overt attention to the CS during conditioning we refer to empirical data and theoretical consideration (Gottlieb, 2012; LePelley et al., 2016; Pearce & Mackintosh, 2010) indicating that attention in our experiment was not just driven by conditioned fear but rather corresponded to the *informational value* of the CS as well (their *predictiveness*). For example, in the healthy control group participants allocated more attention to the safety cue than the contextual cues despite the fact that the latter must have acquired at least some excitatory fear association as part of the aversive conditioning context. In our model the informational value of a CS is encoded in its absolute associative weight and healthy controls from this perspective attended more to the safety cue than the contextual cue because the inhibitory strength of the former exceeds the excitatory strength of the latter (see Figure S3). As detailed above, when the CS- lacks such inhibitory strength because of impaired safety learning in the chronic pain group this attentional preference disappeared and participants with chronic pain equally distributed attention between the (less) inhibitory CS- and the excitatory context.

We have inferred the hypothesis of stronger safety learning in the control group from more attention to the safety signal while our experiment did not include any separate and more direct

test for the acquisition of conditioned inhibition after the conditioning stage (instead we
534 administered a test for automatic attentional capture, see below). We recognize this as an
important task for future research. Specifically, more attention to the safety signal in pain free
536 controls should correspond to more conditioned inhibition as indexed by tests of summation and
retardation (Rescorla, 1969; Williams et al., 1992) as well as an increase in direct measures of
538 positive emotion (see above).

Our original exploratory hypothesis about the search task after conditioning concerned the
540 possibility that during this task differential attention to the CS may be unmasked after attention
was equally high to all CS during fear conditioning because of sensitization processes. As
542 detailed above the premise to this hypothesis was wrong because participants in the chronic pain
group exhibited differences in attention to the CS+ and CS- already during fear conditioning.
544 Accordingly, the results of the search task indicated that automatic attentional capture by the
conditioned cues was intact in *both* groups when (a) shocks were no longer administered, (b)
546 conditioned cues now were task-irrelevant distractors, and (c) any attention to these distractors
was detrimental to performing the task (Anderson, 2015; Schmidt et al., 2015a, 2015b; Wentura
548 et al., 2014). The fact that participants with chronic pain exhibited a modulation of attentional
capture by the previous CS-US contingencies (with $L < P < H$) underline the results from the
550 conditioning stage that while patients did not exhibit differential fear they did exhibit differential
attention. In principle, the observed pattern in the search task is in accord with our account
552 above, however, unfortunately we did not include the contextual stimuli in the search task to
examine how the complete pattern of attentional preference during conditioning would transfer
554 to a pattern of attentional capture in the search task. Specifically, introducing contextual cues

from the conditioning stage should have elicited the same amount of interference than the safety
cue in chronic pain patients. Future research should revisit this question.

Conclusion

While the central role of classical conditioning for the acquisition and preservation of chronic pain has been highlighted previously (Meulders, 2019, 2020; Vlaeyen, 2015) learned attention so far has not been part of that picture. In contrast, our fear conditioning data indicate that relative inattention to the safety signal was the most prominent characteristics in participants with chronic pain. As a possible origin of this *learned inattention*, our theoretical exploration suggests that individuals suffering from chronic pain may have exhibited a weaker encoding of US omission to act a teaching signal for inhibitory learning in CS- trials. With respect to therapeutic implications, this suggests that patients may specifically profit from attempts to improve inhibitory learning during exposure treatments as described for the anxiety disorders (Craske et al., 2014). For example, fear conditioning studies have shown that increasing positive affect during extinction (another form of inhibitory learning) attenuates the reacquisition and reinstatement of fear (Zbozinek & Craske, 2017; Zbozinek et al., 2015) where a boost in positive effect may well relate to a stronger teaching signal to encode US omissions as described in the present article.

Acknowledgements

The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project 222641018 – SFB/TRR 135 TP B4 and project 290878970 – GRK 2271, project 3.

582

Conflict of interest

The authors have no competing interests to declare.

References

- 586 Anderson, B. A. (2015). The attention habit: How reward learning shapes attentional
selection. *Annals of the new York Academy of Sciences*, 1369(1), 24–39.
588 <https://doi.org/10.1111/nyas.12957>
- Baas, J. M. P., Van Ooijen, L., Goudriaan, A., & Kenemans, J. L. (2008). Failure to condition
590 to a cue is associated with sustained contextual fear. *Acta psychologica*, 127(3), 581–592.
- Christianson, J. P., Fernando, A. B. P., Kazama, A. M., Jovanovic, T., Ostroff, L. E., &
592 Sangha, S. (2012). Inhibition of Fear by Learned Safety Signals: A Mini-Symposium
Review. *Journal of Neuroscience*, 32(41), 14118–14124.
594 <https://doi.org/10.1523/JNEUROSCI.3340-12.2012>
- Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing
596 exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, 58,
10–23.
- Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., Hout, M. A., &
Baas, J. M. (2015). Updated meta-analysis of classical fear conditioning in the anxiety
600 disorders. *Depression and anxiety*, 32(4), 239–253
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical
602 power analysis program for the social, behavioral, and biomedical sciences. *Behavior
Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- 604 Ganguly, K., & Kleinfeld, D. (2004). Goal-directed whisking increases phase-locking
between vibrissa movement and electrical activity in primary sensory cortex in rat.
606 *Proceedings of the National Academy of Sciences*, 101(33), 12348–12353.
- Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron*, 76(2), 281–295.

- 608 Green, L., Luck, C., & Lipp, O. (2018). The Effect of Us Intensity on Relief Learning and
Contrast Effects in Backward Conditioning. *Psychophysiology*, 55, S126–S126.
- 610 Grillon, C. (2002). Startle reactivity and anxiety disorders: Aversive conditioning, context,
and neurobiology. *Biological Psychiatry*, 52(10), 958–975.
- 612 Harvie, D. S., Moseley, G. L., Hillier, S. L., & Meulders, A. (2017). Classical conditioning
differences associated with chronic pain: A systematic review. *The Journal of Pain*, 18(8),
614 889–898.
- Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom
616 from sample data in the randomized block and split-plot designs. *Journal of Educational
Statistics*, 1, 69–82. <https://doi.org/doi:10.2307/1164736>
- 618 Icenhour, A., Langhorst, J., Benson, S., Schlamann, M., Hampel, S., Engler, H., Forsting, M.,
& Elsenbruch, S. (2015). Neural circuitry of abdominal pain-related fear learning and
620 reinstatement in irritable bowel syndrome. *Neurogastroenterology & Motility*, 27(1), 114–
127.
- 622 Jenewein, J., Moergeli, H., Sprott, H., Honegger, D., Brunner, L., Ettlin, D., Grillon, C.,
Bloch, K., Bruegger, M., Schwegler, K., Schumacher, S., & Hasler, G. (2013). Fear-
624 learning deficits in subjects with fibromyalgia syndrome? *European Journal of Pain*,
17(9), 1374–1384.
- 626 Kimmel, H., King, J., Hudy, J., & Gardner, K. (1980). A Mutual Inductance Shocker.
Behavior Research Methods & Instrumentation, 12(6), 605–606.
628 <https://doi.org/10.3758/BF03201851>
- Klinger, R., Matter, N., Kothe, R., Dahme, B., Hofmann, U. G., & Krug, F. (2010).
630 Unconditioned and conditioned muscular responses in patients with chronic back pain and

chronic tension-type headaches and in healthy controls. *Pain*, 150(1), 66–74.

<https://doi.org/10.1016/j.pain.2010.03.036>

Koenig, S. (2010). A Software Package for the Analysis of Eye Movement Trajectories with MATLAB. In *Modulation of Saccadic Curvature by Spatial Memory and Associative Learning (Doctoral dissertation, Philipps-Universität Marburg)*. <http://archiv.ub.uni-marburg.de/diss/z2010/0636/>

Koenig, S., Kadel, H., Uengoer, M., Schubö, A., & Lachnit, H. (2017). Reward Draws the Eye, Uncertainty Holds the Eye: Associative Learning Modulates Distractor Interference in Visual Search. *Frontiers in Behavioral Neuroscience*, 11, 128.

<https://doi.org/10.3389/fnbeh.2017.00128>

Koenig, S., Uengoer, M., & Lachnit, H. (2017). Attentional bias for uncertain cues of shock in human fear conditioning: Evidence for attentional learning theory. *Frontiers in Human Neuroscience*, 11(266). <https://doi.org/10.3389/fnhum.2017.00266>

Koenig, S., Uengoer, M., & Lachnit, H. (2018). Pupil dilation indicates the coding of past prediction errors: Evidence for attentional learning theory. *Psychophysiology*, 55(4), e13020.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology. B, Comparative And Physiological Psychology*, 57(3), 193–243.

<https://doi.org/10.1080/02724990344000141>

Le Pelley, Mike E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016).

Attention and Associative Learning in Humans: An Integrative Review. *Psychological Bulletin*, 142(10), 1111–1140. <https://doi.org/10.1037/bul0000064>

- 654 Linton, S. J., Melin, L., & Götestam, K. G. (1984). Behavioral analysis of chronic pain and its
management. In *Progress in behavior modification* (Vol. 18, pp. 1–42). Elsevier.
- 656 Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., &
Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis.
658 *Behaviour Research and Therapy*, 43(11), 1391–1424.
<https://doi.org/10.1016/j.brat.2004.10.007>
- 660 Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J.,
Heitland, I., Hermann, A., Kuhn, M., & Kruse, O. (2017). Don't fear 'fear conditioning':
662 Methodological considerations for the design and analysis of studies on human fear
acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, 77,
664 247–285.
- Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A
666 proposal for standardization. *Psychophysiology*, 8(5), 656–672.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with
668 reinforcement. *Psychological-Review*, 82(4), 276–298.
- Madden, V. J., & Moseley, G. L. (2016). Do clinicians think that pain can be a classically
670 conditioned response to a non-noxious stimulus? *Manual therapy*, 22, 165–173.
- Meulders, A. (2019). From fear of movement-related pain and avoidance to chronic pain
672 disability: A state-of-the-art review. *Current Opinion in Behavioral Sciences*, 26, 130–136.
<https://doi.org/10.1016/j.cobeha.2018.12.007>
- 674 Meulders, A. (2020). Fear in the context of pain: Lessons learned from 100 years of fear
conditioning research. *Behaviour research and therapy*, 131, 103635.

- 676 Meulders, A., Harvie, D. S., Bowering, J. K., Caragianis, S., Vlaeyen, J. W., & Moseley, G.
L. (2014). Contingency learning deficits and generalization in chronic unilateral hand pain
678 patients. *The Journal of Pain*, 15(10), 1046–1056.
- Meulders, A., Jans, A., & Vlaeyen, J. W. (2015). Differences in pain-related fear acquisition
680 and generalization: An experimental study comparing patients with fibromyalgia and
healthy controls. *Pain*, 156(1), 108–122.
- 682 Meulders, A., Meulders, M., Stouten, I., De Bie, J., & Vlaeyen, J. W. (2017). Extinction of
fear generalization: A comparison between fibromyalgia patients and healthy control
684 participants. *The Journal of Pain*, 18(1), 79–95.
- Moseley, G. L., & Vlaeyen, J. W. (2015). Beyond nociception: The imprecision hypothesis of
686 chronic pain. *Pain*, 156(1), 35–38.
- Nagel, B., Gerbershagen, H. U., Lindena, G., & Pfingsten, M. (2002). Development and
688 evaluation of the multidimensional German pain questionnaire. *Schmerz (Berlin,
Germany)*, 16(4), 263–270.
- 690 Notebaert, L., Crombez, G., Van Damme, S., De Houwer, J., & Theeuwes, J. (2011). Signals
of Threat Do Not Capture, but Prioritize, Attention: A Conditioning Approach. *Emotion*,
692 11(1), 81–89. <https://doi.org/10.1037/a0021286>
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the
694 effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87(6),
532–552. <https://doi.org/10.1037/0033-295X.87.6.532>
- 696 Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible
integration. In C. Mitchell and M.E. LePelley (Eds.) *Attention and Learning*, Oxford:
698 Oxford University Press. (pp. 11–39).

Pollak, D. D., Monje, F. J., Zuckerman, L., Denny, C. A., Drew, M. R., & Kandel, E. R.

(2008). An animal model of a behavioral intervention for depression. *Neuron*, 60(1), 149–161.

Pollard, C. A. (1984). Preliminary validity study of the pain disability index. *Perceptual and Motor Skills*, 59, 974. <http://dx.doi.org/10.2466/pms.1984.59.3.974>

Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Frontiers in Neuroscience*, 5, 115. <https://doi.org/10.3389/fnins.2011.0011>

R Core Team. (2020). R: A language and environment for statistical computing.

<https://www.R-project.org/>

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

Rescorla, R.-A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, 72(2), 77–94. <https://doi.org/10.1037/h0027760>

Rogan, M. T., Leon, K. S., Perez, D. L., & Kandel, E. R. (2005). Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. *Neuron*, 46(2), 309–320.

Schmidt, L. J., Belopolsky, A. V., & Theeuwes, J. (2015a). Attentional capture by signals of threat. *Cognition and Emotion*, 29(4), 687–694.

<https://doi.org/10.1080/02699931.2014.924484>

Schmidt, L. J., Belopolsky, A. V., & Theeuwes, J. (2015b). Potential threat attracts attention and interferes with voluntary saccades. *Emotion*, 15(3), 329–338.

<https://doi.org/10.1037/emo0000041>

- 722 Schneider, C., Palomba, D., & Flor, H. (2004). Pavlovian conditioning of muscular responses
in chronic pain patients: Central and peripheral correlates. *Pain, 112*(3), 239–247.
724 <https://doi.org/10.1016/j.pain.2004.08.025>
- Van Damme, S., Crombez, G., & Eccleston, C. (2004). The anticipation of pain modulates
726 spatial attention: Evidence for pain-specificity in high-pain catastrophizers. *Pain, 111*(3),
392–399. <https://doi.org/10.1016/j.pain.2004.07.022>
- 728 Vansteenwegen, D., Iberico, C., Vervliet, B., Marescau, V., & Hermans, D. (2008).
Contextual fear induced by unpredictability in a human fear conditioning preparation is
730 related to the chronic expectation of a threatening US. *Biological psychology, 77*(1), 39–
46.
- 732 Vlaeyen, J. W. (2015). Learning to predict and control harmful events: Chronic pain and
conditioning. *Pain, 156*, 86–93.
- 734 Vlaeyen, J. W., & Linton, S. J. (2012). Fear-avoidance model of chronic musculoskeletal
pain: 12 years on. *Pain, 153*(6), 1144–1147.
- 736 Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Applications
of a theory. In R. A. Boakes & M. S. Haliday (Eds.), *Inhibition and learning* (pp. 301–
738 336). Academic Press.
- Wang, L., Yu, H., & Zhou, X. (2013). Interaction between value and perceptual salience in
740 value-driven attentional capture. *Journal of Vision, 13*(3). <https://doi.org/10.1167/13.3>.
- Wentura, D., Mueller, P., & Rothermund, K. (2014). Attentional capture by evaluative
742 stimuli: Gain- and loss-connoting colors boost the additional-singleton effect. *Psychonomic
Bulletin & Review, 21*(3), 701–707. <https://doi.org/10.3758/s13423-013-0531-z>

- 744 Williams, D. A., Overmier, J. B., & LoLordo, V. M. (1992). A reevaluation of Rescorla's
early dictums about Pavlovian conditioned inhibition. *Psychological Bulletin*, 111(2), 275–
746 290.
- Yalcin, I., & Barrot, M. (2014). The anxiodepressive comorbidity in chronic pain. *Current*
748 *Opinion in Anesthesiology*, 27(5), 520–527.
- Zaman, J., Vlaeyen, J. W., Van Oudenhove, L., Wiech, K., & Van Diest, I. (2015).
750 Associative fear learning and perceptual discrimination: A perceptual pathway in the
development of chronic pain. *Neuroscience & Biobehavioral Reviews*, 51, 118–125.
- 752 Zbozinek, T. D., & Craske, M. G. (2017). The role of positive affect in enhancing extinction
learning and exposure therapy for anxiety disorders. *Journal of Experimental*
754 *Psychopathology*, 8(1), 13–39.
- Zbozinek, T. D., Holmes, E. A., & Craske, M. G. (2015). The effect of positive mood
756 induction on reducing reinstatement fear: Relevance for long term outcomes of exposure
therapy. *Behaviour research and therapy*, 71, 65–75.

758

TablesTable 1. *Demographic characteristics*

Variables	Experimental group (N=25)	Control group (N=25)	Statistics	
			<i>Test</i>	<i>p</i>
Age in years, <i>M (SD)</i>	38.8 (15.31)	36.24 (15.35)	$t(48) = -0.059$	.558
Sex, <i>N (%)</i>			$\chi^2(1) = 0.00$	1
Female	15 (60%)	15 (60%)		
Male	10 (40%)	10 (40%)		
Educational level, <i>N (%)</i>			$\chi^2(2) = 1.064$	.587
Low	2 (8%)	1 (4%)		
Middle	5 (20%)	3 (12%)		
High	18 (72%)	21 (84%)		
Employment status, <i>N (%)</i>			$\chi^2(3) = 1.573$	.665
Employed	22 (88%)	23 (92%)		
Disability pension	1 (4%)	0 (0%)		
Retirement pension	2 (8%)	2 (8%)		
Hospital Anxiety and Depres- sion Scale) (HADS) <i>M (SD)</i>				
Depression	7.17 (4.73)	3.92 (4.2)	$t(47) = -2.54$	.014*
Anxiety	7.79 (4.02)	4.52 (2.22)	$t(47) = -3.55$	.001**

760

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Figures

Figure 1. Experimental design: (a) Fear conditioning trial. Bolt symbolizes shock (or no shock) and is not shown to the participants (b) Search trial.

768

772

Figure 2. Pupil response elicited by the search display: (a) Pupil change over trial time as percent of initial value, (b) average across 5-sec CS interval.

Figure 3. Skin conductance response in anticipation of shock during fear conditioning: (a)

Standardized changes (z) in skin conductance level over trial time. (b) Maximal response magnitude in the cue interval.

780

782 **Figure 4.** Fixation probability over trial time for target (tar: L, H, P), distractor (dis: X, Y,
 783 Z), central fixation cross (central) and other gray distractor stimuli. Shaded area gives differences
 784 in total dwell time.

Figure 5. Total dwell time (sec) on the relevant cue (L, P, H) and the irrelevant distractor

788 (X, Y, Z).

790

Figure 6. Frequency of search trials in which the first fixation was on the color distractor

794 instead of the diamond shape target.

796 **Figure 7.** Predicted group differences in attention (top-panel; softmax normalized absolute
 weights) match the pattern of results observed for fixation probability (Figure 4) and total dwell
 798 time (Figure 5). Predicted group differences in conditioned fear (bottom panel; summed weights
 LX, PX, HX) match the pattern of results observed for pupil size (Figure 2) and skin
 800 conductance (Figure 3).

**Supplement to “An attentional perspective on differential fear conditioning in chronic
pain: The informational value of safety cues.”.**

Stephan Koenig, Karoline Körfer,

Harald Lachnit, & Julia Anna Glombiewski

Changes in pupil size during fear conditioning

Figure S1 depicts changes in pupil dilation across three successive blocks of fear conditioning. The data seem to indicate a general habituation process with overall decrements in pupil dilation across blocks. Differential responding $H > L$ in the control group is intact across all blocks but impaired in all blocks in the chronic pain group. F -test in a general linear model with factors Group (Pain, Control), CS (L, P, H), and Block (1, 2, 3) were: Group, $F < 1$, CS, $F(2, 94) = 9.772, p < .001, \eta^2_p = .169$, and Block, $F(2, 79) = 14.774, p < .001, \eta^2_p = .235$, with interactions Group x CS, $F(2, 94) = 3.209, p = .046, \eta^2_p = .062$, Group x Block, $F < 1$, CS x Block, $F(3, 168) = 1.187, p = .317, \eta^2_p = .024$, and Group x CS x Block $F(3, 168) = 1.137, p = .338, \eta^2_p = .023$.

22

Figure S1. Pupil response elicited by the search display across three successive blocks of fear conditioning. For three different color CS shock expectancy was (L)ow [CS never followed by shock], (P)artial [CS sometimes followed by shock], and (H)igh [CS always followed by shock].

28

30 **Changes in skin conductance during fear conditioning**

Figure S2 depicts changes in the skin conductance response (SCR) across three successive
 32 blocks of fear conditioning. The data again seem to indicate a general habituation process with
 overall decrements in SCR across blocks. In contrast to pupil size however, this decrement seems
 34 to have more extensive affects on differential responding even in the control group which seem
 to exhibit differential responding at the start but not at the end of conditioning. Differential
 36 responding $H > L$ in the chronic pain group again seems entirely absent in the chronic pain
 group. F -tests in a general linear model revealed main effects of Group, $F(1, 47) = 8.545$, $\eta^2_p =$
 38 0.15385 , $p = .005$, CS, $F(1, 47) = 9.935$, $\eta^2_p = .174$, $p = .003$, and Block, $F(2, 79) = 25.8688$, η^2_p
 $= .355$, $p < .001$, with interactions, Group x CS, $F(1, 47) = 3.782$, $\eta^2_p = .0744$, $p = .057$, Group
 40 x Block, $F(2, 79) = 1.935$, $\eta^2_p = .0395$, $p = .157$, CS x Block, $F(2, 86) = 1.022$, $\eta^2_p = .021$, p
 $= .358$, and Group x CS x Block, $F < 1$.

44

Figure S2. Skin conductance elicited by the search display across three successive blocks

46 of fear conditioning.

48

50

Computational explorations of attentional learning in chronic pain patients

Our reinforcement learning model assumes the most essential form of a two-layer network where different input nodes code for the different CS (H, P, L, X, Y, Z), a single output node codes for US strength, and adjustable associative weights connect inputs and output to track the observed CS-US contingencies. Learning is driven by the standard delta-rule also used by the model of Rescorla and Wagner (1972; RW-model)

$$\Delta V_i = a_i \beta (\lambda_{US} - \Sigma V_j) \quad (1)$$

where after each learning episode the change of the associative weight V_i , connecting cue i with the US, is driven by the difference between the observed US strength λ_{US} (teaching signal) minus the combined prediction of the US by all CS present on that trial ΣV_j . In the original RW-model model this error-term (observed minus predicted US) is scaled by a US-related learning rate parameter β (often assumed to be higher for positive errors than for negative errors so that excitatory learning is faster than inhibitory learning, $\beta_{ex} > \beta_{in}$; Harris et al., 2016; Rescorla, 2002) and a CS-related learning rate parameter a_i . While a_i in the RW-model corresponds to the bottom-up salience of the cue only and is assumed to be constant across learning episodes, we take the perspective of most recent attentional learning theories that this bottom-up salience may be modulated by learned attention to the cue. Specifically, we assume a hybrid mechanism that allows for an attentional amplification of predictive cues (Mackintosh, 1975) as well as uncertain cues (Pearce and Hall, 1980; Pearce, Kaye, and Hall, 1982)

$$a_i = \alpha_i + \omega_1 |V_i| + \omega_2 \varepsilon_i \quad (2)$$

where α_i corresponds to the bottom-up salience of CS i , $|V_i|$ corresponds to the absolute value of the associative weight given by Equation (1) and implements attention to predictive

cues, ε_i corresponds to a learned bias to model attention to uncertain CS, and ω_1 and ω_2 are

weights to specify the relative contribution of both attentional mechanisms.

The attentional mechanisms in Equation (2) are straightforward. To capture the notion of *attention to predictive CS* (Mackintosh, 1975; Kruschke, 2001; LePelley, 2004), the model follows suggestions of Esber and Haselgrove (2011) and LePelley (2016) that the predictiveness of a CS is most parsimoniously encoded in the absolute value of its associative weight. Including $|V_i|$ in the combined attention bias ensures that the agent can learn to attend to a CS+ (positive weight) as well as CS- (negative weight) but also can learn to ignore an irrelevant CS (zero weight). To capture the notion of *attention to uncertain CS* (Pearce and Hall, 1980) the uncertainty of a cue is tracked with equation

$$\Delta \varepsilon_i = \gamma (|\lambda_{US} - \Sigma V_j| - \varepsilon_i) \quad (3)$$

where $|\lambda_{US} - a_{US}|$ is the absolute value of the prediction error that also drives learning in Equation 1 and γ is an attentional learning rate parameter. Equation (3) specifies that after each learning episode ε_i is driven towards the absolute error observed in that trial. This weight ensures a boost of attention to CS at the outset of learning (when errors are large) as well as continued attention to partially reinforced CS that consistently lead to some residual prediction error at asymptote. Equation (3) is equivalent to attentional learning as suggested in Pearce, Kaye and Hall (1982; the authors give a state equation instead of a change equation; also see LePelley, 2004). The mechanism of attention to uncertain cues makes a negligible contribution in the current context and our major focus is on attention to predictive cues as reflected in relative weights $\omega_1 > \omega_2$ in Equation 2 as detailed below.

Figure S3. Simulation of an error-driven network. Upper panel show learned weights that correspond to the strength of the shock association. The absolute value of a weight (distance from 0) corresponds to the informational value of a cue manifesting as attention. The lower panels depict the average prediction error associated with a cue, where error is high for cues at the outset of training as well as for partially reinforced cues at asymptote. From left to right panel depict an increase in the strength of the teaching signal encoding US omissions in CS- trials.

Application of the delta-rule in Equation (1) drives the combined prediction ΣV_j towards the teaching signal λ and learning reaches its asymptote when the US is perfectly predicted and the error becomes zero. If we want to train the network about “how much” US to expect we might apply teaching values of $\lambda_{US} = 1$ and $\lambda_{noUS} = 0$ to teach the weights to predict the physical presence and absence of electric shock, respectively. The top-left panel of Figure S3 details this case with parameters $\lambda_{US} = 1$, $\lambda_{noUS} = 0$, $\alpha = .3$ (for all cues), $\beta_{ex} = .3$, $\beta_{in} = .1$, $\omega_1 = .3$, $\omega_2 = .1$, and

$\epsilon_i = .1$. Training the network with the discrimination from our experiment results in excitatory
 108 strength $H > P > X$ asymptotically (Y and Z are equivalent to X and not shown). Irrelevant
 contextual cues X thus are predicted to acquire some excitatory strength (from trials HX+ and
 110 PX+) and it is this contextual fear that is supposed to drive inhibitory learning when combined
 weights $V_L + V_X > 0$ predict more US than actually occurs in LX- trials (where the teaching
 112 signal is $\lambda_{US} = 0$). The negative error $\lambda_{US} - (V_L + V_X) < 0$ causes decrements in associative weights
 that result in a negative (i. e. inhibitory) association for cue L asymptotically that equals the
 114 strength of the context association.

The top-left panel of Figure 7 exemplifies a prominent “standard” explanation of
 116 conditioned inhibition (Rescorla, 1969; Wagner & Rescorla, 1972; Williams et al., 1992) where
 teaching signals of $\lambda=1$ and $\lambda=0$ were used to train the network to predict the physical presence
 118 vs. absence of the US respectively. Excitatory associations are predicted to converge to $H > P > X$
 and the safety signal L acquires an inhibitory association that matches the strength of the
 120 excitatory context association X so that their summed prediction correctly converges towards the
 value of $\lambda=0$ signaling no US in trials LX-. The critical feature to notice about this explanation is
 122 that inhibitory learning entirely is a slave process to contextual fear conditioning: The safety cue
 can only acquire so much inhibitory strength to cancel out conditioned excitation of the context
 124 and how much is learned about the CS- thus entirely depends on how much is learned about the
 context.

126 Surprisingly, this type of “context-dependent” safety learning seems entirely sufficient to
 explain attentional allocation in participants with chronic pain if we follow the suggestion that
 128 the informational value of the cues that drives attention might be encoded in the *absolute* value
 of the associative weight (Esber & Haselgrove, 2011; LePelley et al., 2016). From this

perspective, signed weights $L < X < P < H$ correspond to the emotional value of the cues and signal the level of conditioned threat, while absolute weights $L = X < P < H$ correspond to the informational value of the cues and signal how predictive the cues are for the occurrence *or* omission of the US. The equal distribution of attention between cue L and X in participants with chronic pain that we have identified as a potential attentional deficit thus may relate to the simple notion that in these participants the absolute strength of the safety cue's inhibitory association matched the excitatory context association in accord with the standard explanation of conditioned inhibition in the RW-model. The left panels of Figure 7 show model predictions for total-dwell time (top) and conditioned fear (bottom) derived from asymptotic weights in the top-left panel of Figure S3.

Simulations in the mid panel of Figure S3 model conditioning in the healthy control group and are based on the same parameter set as above except for the stronger negative teaching signal $\lambda_{\text{noS}} = -.3$ to code for US omission in CS- trials. In contrast to participants with chronic pain, healthy controls exhibited a clear attentional preference for the safety signal L over contextual cues X. Applying the same assumptions about the correspondence of absolute weight and attention as above, the model predicts that an increase in attention to cue L would of course result from an increase in the strength of the inhibitory weight. The top-mid panel of Figure S3 depicts learning with a “stronger” non-US teaching signal of $\lambda = -.3$ which to some extent resolves the context dependency of safety learning and coincidentally entails cue L with an increase in attention as well as a decrease in conditioned fear as shown in the right panel of Figure 7. The prediction of group differences in Figure 7 agrees with the empirical data pattern. Specifically, participants with chronic pain exhibit (a) weaker differential responding with

respect to conditioned fear [bottom panels], (b) stronger modulation of attention by the CS [top panels], and (c) a selective “impairment” of attention to the safety signal [top left panel].

Please note that – with the increase in negative λ from left to right – there is a decrease in weight for partially reinforced cues but also an increase in associated error that naturally compensates for that loss. This is most evident for the simulation in the right panel of Figure S3 with $\lambda_{US} = 1$, and $\lambda_{US} = -1$ to code for two completely symmetrical outcomes, and $\beta_{ex} = .2$, $\beta_{in} = .2$ to assume symmetrical learning rates accordingly. The resultant attention biases $H = L > X$ in this case actually correspond to Mackintosh’s (1975) original theory of predictiveness in which attentional learning in each trial is driven by a comparison of cue-specific prediction errors and attention increases for the best available predictor of the outcome and decreases for all other cues. Since the teaching signals for cue P in the right panel of Figure S3 evenly oscillate between 1 and -1 the asymptotic weight of cue P is zero but the error consistently observed after cue P of course is larger, $|\lambda_{US} - \Sigma V_j| = 1$, as for the same oscillation between 1 and 0 in the left panel, $|\lambda_{US} - \Sigma V_j| = .5$. The critical feature here is that an attentional hybrid mechanism (with $\omega_1 > 0$ and $\omega_2 > 0$) will compensate small biases $|V_i|$ with a large bias ϵ_i .

As a caveat to these considerations we would like to point out that shifting non-US λ towards -1 gradually eliminates the context-dependency of conditioned inhibition altogether (see right panel of Figure S3 with symmetric weights for cue L and H and no weight to X). Against this background, we explicitly agree to the original assumption of Wagner and Rescorla (1972) that the context *must* provide some conditioned fear for the learner to perceive the omission of the US as a meaningful event in the first place. There simply is no way to explain how presentations of $LX \rightarrow \text{noUS}$ alone (outside the context of aversive conditioning) could motivate any learning. Rather we assume, that healthy controls – within a moderately excitatory context – were able to

extract a teaching signal from US omission that was to some extent independent from the exact
 176 strength of the excitatory context association.

For the transformation of attentional bias weights a_i into fixation dwell time shown in the
 178 top panels of Figure 7 we use softmax normalization

$$f(a_i) = \exp(\varphi a_i) / \sum \exp(\varphi a) \quad (4)$$

where a corresponds to the learned weight given in Equation (2) and φ is a scaling constant
 182 (set to $\varphi = 1$ for all simulations). Softmax normalization captures the notion that attention is a
 limited resource and restricts summed attention to $\sum f(a_i) = 1$. With respect to the current
 184 experiment this corresponds to the fact that dwell time on cue and distractor are restricted by a
 shared time window. The time spent fixating the cue must decrease dwell-time on the distractor
 186 and vice versa. From this perspective, softmax normalization also captures the notion that
 attention to a stimulus always is context specific. For example, if the same distractor stimulus is
 188 presented in the context of a cue with a high versus low attentional bias weight, the distractor can
 be expected to draw less attention in the former than in the latter case even though its own
 190 attentional bias weight should be the same in both contexts.

To conclude our theoretical explorations, we would like to point out that the RW-model has
 192 been most prominently criticized for the fact that conditioned excitation and conditioned
 inhibition are modeled on the same underlying continuum of associative strength (e. g. ranging
 194 from -1 to 1). In contrast to this assumption it has been shown that extinction learning does in
 fact not seem to correspond to “unlearning” (as the loss of conditioned excitation) but may rather
 196 constitute a form of new learning (the acquisition of an inhibitory association distinct from the
 excitatory association). This problem is addressed by models that include the assumption of a US

node (activated by US presentation) as well as a separate non-US node (activated by US omission) and thus allowing the independent acquisition of excitatory and inhibitory associations in principle. With the assumption of faster increments than decrements in weight for both the CS-US and CS-noUS associations, during extinction the decrement of the CS-US associations is predicted to be smaller than the increase in the CS-noUS association, effectively protecting the CS-US association from complete extinction in accord with empirical findings. With respect to attentional learning, Esber and Haselgrove (2011) have shown that such a model can also account for effects of both attention to predictive cues as well as attention to uncertain cues. The critical aspect with respect to our previous simulation is that such a model, too, must make assumptions about the activation strength of the noUS-node in CS- trials. Importantly, Esber and Haselgrove used a formulation very similar to the RW-model where the noUS-node is activated to the extent that a predicted US did not occur. Here too, learning of the CS-noUS association thus exhibits the same dependency on an excitatory context as shown in the left panel of Figure S3. However, the model in principle would also allow for an activation of the noUS-node that is stronger than the excitatory context association and would thus come to similar results as we have produced here. The applicability to this kind of models illustrates the wider generality of our idea.

References

Esber, G. R., & Haselgrove, M. (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning. *Proc Biol Sci*, 278(1718), 2553–2561. <https://doi.org/10.1098/rspb.2011.0836>

- 220 Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of*
Mathematical Psychology, 45, 812–863.
- 222 Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A
selective review and a hybrid model. *Quarterly Journal of Experimental Psychology. B,*
224 *Comparative And Physiological Psychology*, 57(3), 193–243.
<https://doi.org/10.1080/02724990344000141>
- 226 Le Pelley, Mike E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention
and Associative Learning in Humans: An Integrative Review. *Psychological Bulletin*,
228 142(10), 1111–1140. <https://doi.org/10.1037/bul0000064>
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with
230 reinforcement. *Psychological-Review*, 82(4), 276–298.
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness
232 of conditioned but not of unconditioned stimuli. *Psychological Review*, 87(6), 532–552.
<https://doi.org/10.1037/0033-295X.87.6.532>
- 234 Pearce, J. M., Kaye, H., & Hall, G. (1982). Predictive accuracy and stimulus associability:
Development of a model for Pavlovian learning. In M. Commonsa, R. Herrnstein, & A. R.
236 Wagner (Hrsg.), *Quantitative Analysis of Behavior* (S. 241–255). Cambridge, MA:
Ballinger.
- 238 Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the
effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy
240 (Hrsg.), *Classical conditioning II* (S. 64–99). Appleton Century Crofts.
- Rescorla, R.-A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, 72(2), 77–94.
242 <https://doi.org/10.1037/h0027760>

- Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Applications of a
244 theory. In R. A. Boakes & M. S. Haliday (Hrsg.), *Inhibition and learning* (S. 301–336).
Academic Press.
- 246 Williams, D. A., Overmier, J. B., & LoLordo, V. M. (1992). A reevaluation of Rescorla's early
dictums about Pavlovian conditioned inhibition. *Psychological Bulletin*, 111(2), 275–290.
- 248