Predicting clinical outcomes via human fear conditioning: A narrative review

Sara Scheveneels*[1], Yannick Boddez[2,1], Dirk Hermans[1]

[1]Center for the Psychology of Learning and Experimental Psychopathology, KU Leuven, Leuven, Belgium

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

Funding: This work was funded by a KU Leuven C1 project (C16/19/02) awarded to Dirk Hermans. Yannick Boddez is employed on Ghent University grant BOF16/MET_V/002.

Declarations of interest: none

*Corresponding author:

Sara Scheveneels

Centre for Learning Psychology and Experimental Psychopathology, University of Leuven

Tiensestraat 102 – bus 3712

3000 Leuven, Belgium

+3216323117

Email: Sara.scheveneels@kuleuven.be

Abstract

A common assumption in human fear conditioning research is that findings are informative for the etiology and treatment of clinical anxiety. One way to empirically evaluate the external validity of fear conditioning is by prospective studies. We review available prospective research investigating whether individual performance in fear conditioning predicts individual differences in anxiety levels and exposure-based treatment outcome. We focus on fear extinction, generalization, acquisition, and avoidance.

Results suggest that reduced extinction and broader generalization predict higher anxiety levels. Results with respect to the predictive value of acquisition for anxiety levels are mixed. With regard to predicting exposure-based treatment outcome, some studies do find an association with extinction whereas others do not. The majority of studies does not find an association with acquisition. Evidence on extinction recall is limited and not consistent.

The interpretation of these results requires caution. The number of available studies is limited. It is possible that not all work, in particular studies with only null effects, has found its way to publication. Future research on this topic will benefit from large sample sizes, preregistered hypotheses, full transparency about the conducted analyses and the publication of high-quality studies with null effects.

Keywords: fear conditioning, anxiety disorders, exposure therapy, extinction, external validity

Predicting clinical outcomes via human fear conditioning: A narrative review

Fear Conditioning

The fear conditioning procedure has widely been used as a laboratory model to study the etiology and treatment of fear and anxiety (e.g., Beckers, Krypotos, Boddez, Effting, & Kindt, 2013; De Houwer, 2020; Mineka & Zinbarg, 2006). The first laboratory demonstration of fear conditioning goes back to the study of "Little Albert" by Watson and Rayner (1920). The 11-month old Albert was presented with a white rat, which initially did not evoke fear. It was only after pairing the rat with a loud noise that Albert started to react fearfully to the rat (Hermans, Boddez, & Vervliet, 2019). This procedure is fairly similar to the basic procedure that fear conditioning researchers currently use in the laboratory (Lonsdorf et al., 2017). In fear acquisition, a stimulus (i.e., the *conditional stimulus*; CS+) that initially does not evoke fear, is paired with an aversive stimulus (i.e., the *unconditional stimulus*; US). After pairing the CS+ with the US, the CS typically starts to elicit *conditional responding* indicative of fear and anxiety. A second CS (i.e., CS-) is often included with the aim to provide a baseline measurement (i.e., to control for changes in responding that are not due to stimulus pairings). This CS- is typically presented equally often as the CS+, but is never followed by the US. A variety of stimuli can be used as CSs (e.g., geometrical shapes, [fearful] faces, spiders) and as US (e.g., electric shocks, aversive pictures or movie clips, loud noises or human screams). Moreover, the conditioning procedure allows to study multiple indices of fear and anxiety, including verbal ratings (e.g., US-expectancy, subjective fear), physiological indices (e.g., skin-conductance, fear-potentiated startle), and brain activity (e.g., fMRI, electroencephalography).

Fear conditioning allows for the study of fear in highly controlled experimental designs, thereby contributing to an excellent internal validity. Decades of fear conditioning research in

rodents and humans contributed to a solid knowledge base (i.e., Craske, Hermans, & Vansteenwegen, 2006).

Is fear conditioning a model for clinical fears and anxiety and its treatment?

Fear conditioning has been used to study fear as a basic emotion (LeDoux, 2000) and to study how fear may contribute to psychological suffering (Mineka & Zinbarg, 2006). Typically, fear and anxiety responses are associated with higher levels of suffering when they (1) are intense, (2) are persistent over time, (3) are triggered by a large set of stimuli or situations and (4) lead to interference with daily functioning. For example, someone who has been involved in a biting incident with a dog might continue to react with intense fear to all types of dogs, which interferes with his daily life. Four candidate processes can be proposed to explain these features of clinical anxiety: (1) strong acquisition; (2) slow extinction; (3) broad generalization; and (4) extensive avoidance. We will now discuss further how these candidate processes might be relevant in clinical anxiety and its treatment and how they can be modeled in fear conditioning.

A first candidate process, that can account for the intense character of clinical anxiety, is **strong acquisition**. After a conditioning experience in real life (e.g., being bitten by a dog), some individuals might exhibit more intense anxiety (e.g., to dogs) than others. This could be explained by stronger fear acquisition in these individuals. As we discussed earlier, fear acquisition in the laboratory entails pairings of a CS with a US. Existing studies indicate that there are differences between anxious and non-anxious individuals in the acquisition of fear in the laboratory (e.g., Duits et al., 2015; Lissek et al., 2005[1]; but see Pöhlchen et al., 2020).

A second candidate process that is presumed to be involved in clinical anxiety is **slow extinction**. After a conditioning experience in real-life, natural decreases in fear may occur as

[1] Notably, Lissek et al. (2005) found patient-control differences primarily in simple conditioning paradigms with a single CS rather than in differential fear conditioning.

a result of spontaneous confrontations with the feared stimulus (e.g., dogs) in the absence of the aversive consequences (e.g., being bitten). However, in individuals who show slower extinction, this natural decrease in fear will be delayed or prevented. As a consequence, these individuals will continue to react fearfully to conditional stimuli despite the experience that the aversive outcome does not occur. These individuals may be more vulnerable to develop an anxiety disorder after an aversive experience because of sustained fear reactions and limited impact of corrective (safety) information. In a fear conditioning procedure, extinction is modelled by adding trials in which the (now fear-inducing) CS+ is repeatedly presented in the absence of the US (Hermans, Craske, Mineka, & Lovibond, 2006). Interestingly, differences between anxious and non-anxious individuals have been found in the extinction of conditioned fear (e.g., Duits et al., 2015, but see Pöhlchen et al., 2020). In addition, the extinction procedure is typically seen as the laboratory analog of exposure treatment, which is the psychological treatment of choice for clinical anxiety (Öst, Havnen, Hansen, & Kvale, 2015; Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008; but also see Scheveneels, Boddez, Vervliet, & Hermans, 2016). Similar to lab-based extinction, exposure treatment involves the repeated confrontation with the fear-eliciting stimulus or situation without occurrence of the aversive outcome with the aim to reduce clinical fear and anxiety (McNally, 2007).

Third, **broad generalization** has been proposed as a candidate process involved in clinical anxiety (Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015). Clinical fears and anxiety are rarely limited to the specific stimuli and situations involved in the conditioning experience. For example, exhibiting fear to the specific dog involved in a biting incident can be considered adaptive. However, it can become more debilitating if also other dogs, pictures of dogs, the sound of barking, cats, etc. evoke fear responding. Hence, individuals who are more prone to generalize might exhibit fear to a broader set of stimuli and situations. In a fear conditioning procedure, generalization of fear can be investigated by inspecting fear

responding towards other stimuli than the original CS+. To this end, a generalization test phase – including generalization stimuli – can be added to the basic fear conditioning procedure. Lissek et al. (2008), for instance, used small and large rings that served respectively as CS+ and CS- (counterbalanced) during acquisition. After acquisition, fear generalization is tested by examining responding to rings of sizes varying between the CS+ and CS-. Typically, a generalization gradient is observed with higher fear responding towards stimuli that show more resemblance to the CS+ and diminishing fear responding towards stimuli more similar to the CS- (e.g., Vervliet, Vansteenwegen, & Eelen, 2006). It is interesting to note that several studies indicate that anxiety patients and non-anxious controls differ in generalization in the laboratory (e.g., Lissek et al., 2010).

A fourth candidate process involved in clinical anxiety is the extent of **avoidance**. Avoidance can be defined as overt or covert behavior directed at decreasing the probability of being confronted with an aversive stimulus (De Houwer & Hughes, 2020). It often interferes with other goals and therefore can have a severe impact on individuals' daily functioning (e.g., loss of social activities; Salters-Pedneault, Tull, & Roemer, 2004). Avoidance can additionally contribute to the persistence of suffering. Confrontation with the fear-eliciting stimuli is a prerequisite for the extinction of fear responding. Avoiding the fear-eliciting stimuli takes away the opportunity to experience that the expected aversive consequences do not occur, potentially resulting in retained fear responding (Lovibond, Mitchell, Minard, Brady, & Menzies, 2009). Moreover, empirical evidence shows that avoidance behavior can even induce increased fear responding (Vervliet & Indekeu, 2015). To test avoidance in the laboratory, the basic fear conditioning procedure can be extended with an instrumental phase. In this phase, participants are given the opportunity to cancel the US by performing (active avoidance) or withholding (passive avoidance) a response during presentation of the CS. To model interference with other life goals as seen in anxiety disorders, costs (e.g., loss of points)

associated with performing the avoidance response can be included in the standard procedure[2]. Interestingly, differences in the nature of conditioned avoidance have been found between anxiety patients and non-anxious controls (Gillan et al., 2011).

In conclusion, individuals who show strong acquisition, slow extinction, broad generalization and extensive avoidance might be more prone to develop clinical anxiety. As such, factors at the level of the individual can be considered a moderator between a conditioning experience in real life and clinical anxiety (Figure 1). The above indicates that these candidate processes involved in clinical anxiety can be modelled in fear conditioning in the laboratory. The crucial question now remains whether fear conditioning is indeed a valid model for studying the onset and treatment of clinical anxiety.

What about the external validity of fear conditioning?

Fear conditioning has a tradition of setting up highly controlled experimental designs in order to exert rigorous control over confounding variables. As an illustration, Lonsdorf and colleagues (2017), in collaboration with representatives from fourteen of the major laboratories working on human fear conditioning, provided a set of guidelines on methodological considerations to even further improve the internal validity of experimental designs in human fear conditioning research. However, these extensive efforts with regard to optimizing the internal validity stand in contrast to the limited efforts that have been made to verify the external validity of fear conditioning research. This is remarkable given that the ultimate goal of a large part of fear conditioning research is to be informative for clinical fears and its treatment (Scheveneels, Boddez, & Hermans, 2019; Vervliet & Boddez, 2020).

There are several ways to evaluate the external validity of the fear conditioning model in an empirical way (Vervliet & Raes, 2013). The current review focusses on evidence from prospective studies that investigate whether individual performance in fear conditioning

[2] For an elaborate evaluation of avoidance paradigms, we refer to Krypotos, Vervliet, and Engelhard (2018).

predict individual differences in clinical anxiety and treatment outcome (Kraemer et al., 1997). We provide an overview of available prospective research on human fear conditioning in predicting 1) (clinical) anxiety levels in real life and 2) the outcome of exposure-based treatment. In line with the candidate processes that are presumed to be involved in clinical anxiety and its treatment, we focus on the predictive value of lab-based fear extinction, generalization, acquisition, and avoidance.

Search strategy

We conducted literature searches in Web of Science, using the following search terms: (1) (predic* AND conditioning AND longitud* AND anxiety); (2) (predict* AND extinction AND anxiety); (3) (predict* AND generalization AND anxiety) and (4) (predict* AND avoid* AND anxiety AND conditioning). In addition, we screened the reference lists of the selected articles. After screening of the abstracts, a total of 13 articles was identified as relevant for this review (Figure 2).

Results

Does human fear conditioning predict anxiety levels in real life?

In this section we describe prospective research investigating whether lab-based fear conditioning can predict levels of (clinical) anxiety in real life. We found a total of five studies on this topic. The results of these studies are further classified according to the specific candidate process that serves as predictor. Three results were found on extinction, one result on generalization, four results on acquisition and no results on avoidance. Notably, some of the studies report on several candidate processes and are discussed under more than one section. An overview of the studies and results can be found in Table 1.

Extinction

A first study by Guthrie and Bryant (2006) investigated whether pre-trauma individual differences in extinction learning predict posttraumatic stress. Seventy firefighters were tested during training (before exposure to a traumatic event) and posttraumatic stress was assessed within 24 months after trauma. Testing consisted of a fear conditioning task and several measures of (baseline) posttraumatic stress. In the initial acquisition phase, a CS+ was repeatedly paired with a shock US, whereas the CS- was presented an equal number of times without the US. In the extinction phase both the CS+ and CS- were presented without the shock US. CSs were colored circles. Corrugator electromyography (EMG) and skin-conductance (SC) were measured throughout the conditioning task. Only participants who were aware of the CS-US contingency after acquisition (N = 45) were included in the analyses. Notably and perhaps in contrast with currently promoted research practices (Lonsdorf et al., 2019), to identify a subset of potential predictors, the authors calculated correlations between the primary outcome measure of posttraumatic stress (i.e., the Impact of Event Scale; IES) and several indices of acquisition and extinction learning in both EMG and SC. The psychophysiological variable that correlated highest with the IES (i.e., differential extinction in corrugator EMG throughout extinction) was included in the further analyses. In addition, a measure to control for the effects of prior traumatic events (i.e., Traumatic Event Questionnaire) was included. Reduced extinction of the conditioned corrugator EMG response (i.e., mean response during CS+ trials minus mean response during CS- trials throughout extinction) predicted higher posttraumatic stress (above prior traumatic events). Notably, differential SC during extinction did not correlate significantly with IES score.

Similar to Guthrie and Bryant (2006), Lommen, Engelhard, Sijbrandij, van den Hout, and Hermans (2013) investigated whether individual differences in extinction learning before deployment in Afghanistan predict subsequent PTSD symptom severity. Their study design aimed to solve some of the limitations of the study by Guthrie and Bryant (2006). They

included a larger sample and controlled for baseline PTSD symptoms and other risk factors such as pre-trauma neuroticism and stressor severity to assess the unique contribution of extinction learning. Two-hundred forty-nine Dutch soldiers completed a fear conditioning task before their 4-month deployment to Afghanistan. In addition, neuroticism and PTSD symptoms were measured. Two months after deployment, PTSD symptom severity was measured again, as well as exposure to stressors during deployment. The fear conditioning task consisted of an acquisition phase in which a CS+ was presented together with an electric shock (US), whereas a CS- was never presented with the shock. This phase was followed by an extinction phase in which both CS+ and CS- were presented without the shock. CSs were pictures of neutral human faces. US-expectancy ratings were collected on a trial-by-trial basis. Thirty-three participants were excluded from the analyses because they did not understand the conditioning task instructions or did not meet the acquisition criterion. Reduced pre-trauma extinction learning (as measured by a difference score between CS+ expectancy ratings on trial 1 and CS+ expectancy ratings on the fourth trial of the in total eight extinction trials) predicted higher PTSD symptoms (as measured by the Posttraumatic Stress Scale) after deployment, even after controlling for baseline PTSD symptoms, neuroticism and stressors during deployment. Notably, the predictive effect of extinction learning on PTSD symptoms was significant when inspecting reduction in the first part of extinction (CS+ on trial 1 minus CS+ on trial 4) but disappeared when looking at later extinction learning (from trial 1 to 6 and beyond). The authors therefore suggest that deficits in extinction learning could be overcome if additional extinction trials are given. This result could imply that vulnerable individuals can overcome their anxiety, but might need additional natural exposure.

In a sample of police and firefighter trainees (N = 211), Orr and colleagues (2012) investigated whether pre-trauma fear extinction (as part of a larger set of predictors) predicts post-traumatic stress symptoms. Participants were assessed at the start of training (prior to the

occurrence of a traumatic event) and were subsequently contacted by email on a bi-monthly basis to assess whether a traumatic event occurred. Ninety-nine participants were invited for post-assessment after they reported a traumatic event. The fear conditioning task was administered before trauma. During acquisition, the CS+ was paired with a shock-US, the CS- was presented without shock. During the subsequent extinction phase, both stimuli were presented in the absence of the shock-US. CS's were a blue and a white circle. SC, heart rate and left corrugator EMG were measured. In addition to fear conditioning, depressive symptoms, IQ, trait anxiety, stressful life events, personality traits and psychophysiological reactivity to loud noises were included as predictors. The primary outcome measures were the Impact of Event Scale (IES) and psychophysiological reactivity (i.e., heart rate, SC and EMG) during personalized, trauma-related script-based imagery. Data analysis focused on (1) relationships between the separate pre-trauma predictors and posttraumatic stress symptoms (measured by the IES and reactivity during imagery, both dichotomized as high versus low), and (2) a stepwise logistic regression including combined pre-trauma predictors of posttraumatic stress symptoms (measured by the IES and reactivity during imagery, both dichotomized as high versus low). With regard to the separate analyses, we focus on the results of the fear conditioning task. Larger differential corrugator EMG during extinction (i.e., mean responses to CS+ trials subtracted by mean responses to the CS- trials during the extinction phase) predicted higher posttraumatic stress as measured with the IES. Logistic regression analysis identified lower IQ, higher depression score and poorer extinction of corrugator EMG (i.e., larger differential corrugator EMG during extinction) as the best predictors of higher posttraumatic stress as measured by the IES[3].

[3] Notably, the effect of corrugator EMG on posttrauma symptoms was the result of higher responding to the CS- than to the CS+ in the low IES group (resulting in a negative differential score).

Based on the three available studies on predicting anxiety levels by lab-based extinction, reduced extinction (either less decline in early extinction or larger differential fear responding throughout extinction) seems to predict higher anxiety.

Generalization

We found only one study investigating whether individual differences in generalization prospectively predict subsequent (subclinical) levels of anxiety. Of note, this study employed a picture-picture paradigm (as also used in studies on evaluative learning; e.g., De Houwer, Baeyens, Vansteenwegen, & Eelen, 2000) and measured subclinical levels of anxiety. In this study by Lenaert et al. (2014), a sample of first-year university students (N = 375) completed a conditioning task and their (subclinical) anxiety levels were assessed six months later. First-year university students were tested because the transition to university can be characterized by a set of stressors (Dyson & Renk, 2006). The conditioning task was similar to the procedure used by Lissek et al. (2008) and consisted of an acquisition phase with small and large circles that served as CS+ and CS- respectively. Aversive pictures of the International Affective Picture System (IAPS) served as US. During acquisition, one circle (CS+) was paired with a US in nine of the 12 trials, the other circle (CS-) was never paired with the US. After acquisition, a generalization test phase followed including presentations of the CS+ and CS- as well as eight generalization stimuli. These generalization stimuli were circles with their size varying on a continuum between the CS+ and CS- (similar to Lissek et al., 2008). US-expectancy ratings were measured on a trial-by-trial basis during the conditioning task. Anxiety was measured at baseline and again after six months by the Depression Anxiety Stress Scales (DASS-21) and State-Trait Anxiety Inventory (STAI-S). The DASS-21 assessed participants' emotional states during the past week. Results showed that the extent of generalization in the US-expectancy ratings (i.e., elevated US-expectancy ratings during the generalization stimuli that resembled the CS- the most) was associated with

higher levels of anxiety on the DASS-21 after six months (after controlling for anxiety levels at baseline). No significant effects were found on the STAI-S. In addition, anxiety (DASS-21) and generalization were not associated at baseline, suggesting that generalization was associated with subsequent increases in anxiety and not merely reflected current anxiety.

Acquisition

Most studies that investigate whether acquisition learning predicts (clinical) anxiety focus primarily on other learning processes such as extinction and generalization. These studies are therefore discussed more elaborately in those sections.

In addition to generalization, Lenaert et al. (2014) examined whether lab-based differential US-expectancies during acquisition predict later (subclinical) anxiety levels. It was found that impaired discrimination learning (i.e., summed CS- US-expectancy ratings on the last three acquisition trials subtracted from summed CS+ US-expectancy ratings on the last three acquisition trials) predicted higher levels of anxiety after 6 months (as measured with the DASS-21) after controlling for baseline anxiety. When analyzing US-expectancy ratings during the CS- and CS+ separately, elevated US-expectancy ratings towards the CS- on the last three trials were found to predict higher anxiety levels six months later (DASS-21). A similar trend was found for decreased US-expectancy ratings during the CS+.

In their sample of police and firefighter trainees, Orr et al. (2012) examined whether, in addition to extinction, indices of lab-based acquisition are predictive for later posttraumatic stress. Results show that a larger mean heart rate response to CS+ trials during acquisition predicted higher posttraumatic stress as measured by psychophysiological reactivity to the script-based imagery.

Although the main focus of the study of Guthrie & Bryant (2006) was on extinction learning, they also report results on acquisition. Differential EMG responding during

acquisition was positively correlated with IES score (as a measure of posttraumatic stress), whereas differential SC during acquisition did not correlate significantly with IES score.

Sijbrandij, Engelhard, Lommen, and Leer (2013) investigated whether individual differences in fear acquisition predict the persistence of PTSD symptoms. The sample was a group of 144 Dutch soldiers who had been deployed in Afghanistan. Two months after their deployment, a fear conditioning task was administered. PTSD symptoms were measured using the Posttraumatic Symptom Scale – Self Report at two and nine months after deployment. In the conditioning task, a geometrical shape (X) was paired with an electric stimulus (US) only if X was presented together with one geometrical shape (AX+), but not if X was presented with a second geometrical shape (BX-). US-expectancy ratings and fear-potentiated startle were measured during this task. It was examined whether performance in this task predicted PTSD symptoms nine months after deployment (controlled for PTSD symptoms at two months post-deployment). In contrast with the findings of Lenaert et al. (2014) impaired discrimination learning (i.e., mean AX+ response minus mean BX- response on the three last trials) in both the startle and US-expectancy measure did not predict the persistence of PTSD symptoms after nine months.

In conclusion, the results on lab-based acquisition as a predictor of anxiety levels in real life are mixed. Some studies indicate that increased US-expectancy ratings to the CS- (i.e., a stimulus signaling safety) and impaired CS+/CS- discrimination learning during acquisition predicts higher anxiety (Lenaert et al., 2014). Other studies find that higher responding to the CS+ and larger CS+/CS- discrimination predict higher anxiety (Guthrie & Bryant, 2006 in EMG; Orr et al. 2012). Yet other studies did not find significant correlations between acquisition learning and anxiety (Guthrie & Bryant, 2006 in SCR; Sijbrandij et al., 2013).

Avoidance

We did not find any studies investigating whether individual differences in avoidance learning during lab-based fear conditioning predict (clinical) anxiety levels.

Does human fear conditioning predict exposure-based treatment outcome?

In this section, we review prospective studies on fear conditioning as a predictor of exposure-based treatment outcome. Our literature search resulted in a total of seven studies on this topic. We classified the results again according to the involved candidate learning processes. All seven studies report results on extinction, six of the studies report results on acquisition and two studies report results on extinction recall. No studies on generalization or avoidance were found. An overview of the studies and results can be found in Table 2.

Extinction

Waters and Pine (2016) examined whether fear extinction predicts response to CBT in 7- to 13-year old children. They included 44 anxious children with different principal diagnoses of anxiety disorders and 33 non-anxious controls. Before treatment, a diagnostic interview and continuous symptom measure were conducted as well as a fear conditioning task. In the acquisition phase of the conditioning task, a CS+ was always followed by a 100 dB tone (i.e. US), whereas a CS- was never followed by the tone. During extinction, both the CS+ and CS- were presented without the US. CSs were colored geometrical shapes. SC was measured on a trial-by-trial basis and subjective valence and arousal ratings were assessed after each phase of the conditioning task. SCRs were averaged over two blocks of two trials for each CS. Subsequently, the anxious children completed a 10-week group-based CBT program including exposure but also other treatment components, such as psychoeducation, relaxation techniques and socials skills training. After treatment, the diagnostic interview and continuous symptom measure were administered again. Based on their diagnostic status after treatment, participants were identified as being a treatment responder (n = 26) or a non-responder (n = 18). Treatment responders did no longer meet the diagnostic criteria of an

anxiety disorder after treatment, whereas non-responders did. Differences between responders and non-responders were found during extinction. Treatment responders and non-anxious controls showed a decline in overall SC (averaged over the CS+ and CS-) from the first to the second block of extinction, whereas non-responders did not show this decline[4]. Moreover, a significant positive correlation was found between the continuous symptom change scores and SCR to the CS+ during the first block of extinction. No other correlations were significant. With regard to the ratings data from pre- to post-extinction, no group differences were found. No significant correlations were found between continuous symptom change and the ratings data. The authors conclude that treatment responders and non-anxious controls show a similar decline in SC during extinction, whereas non-responders failed to show a decline in SC during extinction.

Wannemueller et al. (2018) tested whether fear extinction predicts the outcome of a highly standardized exposure-based treatment. They selected adults fearful of spiders (n = 77), dental surgery (n = 43) and blood, injuries or injections (n = 40) who requested treatment. Prior to exposure, a fear conditioning task was administered in collective test sessions. Two Rorschach figures served as CSs and were projected on a large screen in a lecture hall. The US was a 85 dB scratching noise presented through speakers. During acquisition, the CS+ was paired with the US in 80 percent of the trials. The CS- was never paired with the US. After a break of 40 minutes, the extinction phase followed in which the CS+ and CS- were each presented in the absence of the US. Subjective fear ratings (on a scale from 0 to 10) and US-expectancy ratings (on a dichotomous yes/no scale and on a scale from 0 to 100) were assessed before acquisition, after the fifth acquisition trials (early acquisition), after the last acquisition trial, after the fifth extinction trial (early extinction) and after the last extinction

[4] Notably, in the first block of extinction, overall SC (averaged over CSs) was larger in treatment responders compared to non-responders, while the decline in SC in responders resulted in similar overall SC magnitudes as in non-responders by the end of extinction.

trial. A differential extinction index was calculated by subtracting subjective fear ratings and US-expectancy ratings during early or late extinction from late acquisition scores. After extinction, a one-session standardized exposure was provided in large group settings, including psychoeducation as well as video, pictorial and live exposure exercises. Phobic fear was assessed before fear conditioning, after the exposure treatment and again at 7-month follow up by asking participants how fearful they would be if they imagined seeing their feared object (11-point scale; 0 = not at all, 10 = extremely). No significant correlations were found between differential extinction and fear reductions immediately after treatment or at follow-up.

Geller et al. (2019) examined whether lab-based fear extinction prior to CBT predicts treatment response. In this study the sample consisted of 142 participants between 7 and 17 year with moderate levels of obsessive-compulsive symptoms. The research question was part of a larger randomized-controlled trial comparing CBT with and without d-cycloserine. The 10-session CBT protocol included 6 sessions of exposure and response prevention as well as psychoeducation and cognitive interventions. Prior to CBT, participants completed a fear conditioning task. During acquisition neutral female faces were presented on the screen. One of these faces (i.e., CS+) was followed by a 95-dB scream and fearful facial expression on the screen (i.e., US) in 80 percent of the trials. The other face was never followed by the US. During extinction, both faces were presented without the US. SC was measured throughout the fear conditioning task. Notably, 64 of 142 participants completed at least one block (i.e., two trials) during acquisition and only 38 participants completed the entire fear conditioning task. This reduces the sample and power of this study significantly. Participants were categorized as responders or non-responders based on their post-CBT symptom scores. Notably, the group of non-responders that completed the extinction phase consisted of only 9 participants. Condition (with and without d-cycloserine) was added as a covariate in the

analyses. Results indicate that baseline obsessive-compulsive symptoms were not associated with extinction learning. Differences between responders and non-responders were found in the extinction phase. Responders differentiated between the CS+ and CS- during early extinction and maintained differential responding throughout extinction. Non-responders did not differentiate between the CS+/CS- at the start of extinction and continued to exhibit fear to the CS- throughout extinction.

Forcadell et al. (2017) investigated whether individual differences in extinction are associated with the outcome of an exposure treatment analog. Fifty adults with moderate to strong fear of spiders participated in a 2-day experiment. On day 1, baseline fear of spiders was measured by a questionnaire and participants completed the acquisition and extinction phase of a fear conditioning task (based on Milad, Orr, Pitman, & Rauch, 2005). During acquisition, a picture of a room (context A) containing a lamp that switched on to two different colors (i.e., CS+ and CS-) was presented. If the lamp switched to the CS+ color, it was followed by an electric shock (i.e., US). If the other color (i.e, CS-) was switched on, the US did not occur. During extinction, a picture of another room (context B) was presented containing a lamp that was switched on to the CS+ and CS- colors. Neither the CS+ nor the CS- were followed by the US. US-expectancy ratings, SC and FPS were measured throughout the entire fear conditioning task. Participants also engaged in an exposure treatment analog, which consisted of repeated presentations of the two most fear-eliciting images of spiders each participant selected out of 30 images. The total duration of the exposure was 20 minutes. Outcome measures were subjective ratings on a visual analogue scale (VAS), SC and FPS measured during exposure and post-exposure scores on a fear of spiders questionnaire. Both the percentage of fear reduction after exposure and pre-post differences in VAS, FPS and SC were calculated. The exposure was effective in reducing fear in all outcome measures. Enhanced extinction learning as measured in US-expectancies (i.e. differences between the

CS+ and CS- during the first three extinction trials and operationalized as a percentage) was associated with greater pre-post reductions in the VAS. In a selection of participants who showed acquisition and extinction learning, enhanced differential extinction learning in US-expectancies and FPS during the first three extinction trials predicted greater pre-post reductions in VAS and FPS as well as a larger percentage of fear reduction in FPS.

Raeder et al. (2020) investigated whether conditioning in the laboratory predicts the ability to complete exposure treatment in a predetermined time. The sample consisted of 53 spider phobic individuals who were seeking treatment. Prior to exposure treatment, participants' baseline fear of spiders was assessed using questionnaires and a behavioral approach task. In addition, participants completed a fear conditioning task in virtual reality. During acquisition, the CS+ and CS- (i.e., a high and low frequency tone) were presented in context A. The CS+ was followed by an electric shock (i.e., US) in 60 percent of the trials. The CS- was never paired with the US. After an interval of 10 minutes, an extinction phase followed with both the CS+ and CS- presented without the US in context B. SC was measured on a trial-by-trials basis and valence of the CSs and CS-US contingency ratings were asked retrospectively after each phase of the conditioning task. SCRs were averaged across four trials during extinction (i.e., early and late extinction). Immediately after exposure treatment and at 6-weeks follow-up, exposure outcome was measured using the questionnaires and the behavioral approach task. Participants who were able to complete all exposure steps within the allotted time were classified as completers (n= 29). Other participants were classified as non-completers (n= 24). Completers showed enhanced short- and long-term outcomes of exposure treatment in the subjective fear measures. No differences between completers and non-completers were found (in valence and contingency ratings and in SCR) in lab-based fear extinction.

The studies described above focus on behavioral and psychophysiological measures during fear conditioning. Additionally, we found two studies that focus on brain activation during fear conditioning in predicting treatment response. A first study was conducted by Ball et al. (2016) and examines whether brain activation during fear extinction could predict exposure outcome in public speaking anxiety. Twenty-four adults anxious of public speaking completed a fear conditioning task during fMRI scanning. During acquisition, the CS+ was paired 10 times with a loud scream (i.e., US) and presented 30 times without the US. The CS- was presented 30 times without the US. Extinction immediately followed the acquisition phase and consisted of unreinforced CS+ trials and CS- trials. The CS+ and CS- were two abstract images. After acquisition and extinction, participants rated the CSs on valence and arousal. In a subsequent session the rationale for exposure therapy was presented and participants were instructed to give four speeches of 5 minutes each in front of a video camera, the therapist and two confederates. Self-reported public speaking anxiety and social anxiety were administered at baseline (i.e. before the conditioning task) and again two weeks after the exposure session. Primary hypotheses focused on whether brain activation during the extinction phase predicted exposure outcome. Based on previous studies on brain activation during fear extinction and exposure therapy (e.g., Milad & Quirk, 2002; Whitfield-Gabrieli et al., 2016), it was predicted that greater vmPFC and less amygdala activation during extinction would predict greater anxiety reduction following exposure. For exploratory reasons, it was also tested whether valence and arousal ratings predicted reductions in public speaking anxiety and social anxiety. The results showed that the exposure session in general resulted in reduced public speaking anxiety and social anxiety. Greater vmPFC activation during early extinction, but not amygdala activation, predicted greater reductions in *public speaking anxiety* after exposure. For *social anxiety* – as hypothesized – greater vmPFC activation and less amygdala activation during extinction were associated with greater reductions from pre-

to post-exposure. In addition, less activation in parietal and occipital cortical regions during early extinction predicted greater reductions in *public speaking anxiety*. Less activation in the peri-aquaductal gray (PAG) during early extinction and right anterior insula during late extinction were associated with higher reductions in *social anxiety*. The authors conclude that better extinction learning – as evidenced by less activation in fear-related brain regions – predicts better exposure outcomes. Valence and arousal ratings (i.e., CS+/CS- differentiation after extinction) during extinction were not associated with reductions in *public speaking anxiety*. However, smaller differences between the CS+ and CS- valence ratings after extinction predicted a greater reduction in *social anxiety*.

In a second fMRI study, Hahn et al. (2015) investigated whether they could predict CBT response based on fMRI data measured during a fear conditioning task. Forty-nine patients with panic disorder and agoraphobia underwent fear conditioning during fMRI data acquisition before twelve twice-weekly sessions of CBT focusing on exposure. The fear conditioning task consisted of an acquisition phase in which the CS+ was paired with an unpleasant white noise (i.e., US) in 50 percent of the trials and presentations of a CS- without the US. During subsequent extinction the CSs were never followed by the US. CSs were colored geometrical shapes. A novel multivariate pattern classification approach considering local and whole-brain information was used to predict treatment response in individual patients. Patients were categorized as responders (n = 25) if they reported a reduction of 50 percent or more in self-reported anxiety symptoms from pre- to post-CBT. Results indicate that activity in no single brain region was predictive of CBT response. When integrating regional data from whole brain data, patients could be classified correctly as responders or non-responders with an accuracy of 82 percent. The regions that contributed to the accurate prediction partly overlap with the fear network, including the orbitofrontal cortex and inferior frontal gyrus.

In conclusion, available evidence on the predictive value of individual differences in lab-based extinction learning for exposure-based treatment outcome is mixed. Some studies provide evidence that enhanced extinction learning is associated with better exposure treatment response (Ball et al. 2016; Forcadell et al., 2017; Waters & Pine, 2016). Other studies find an association between extinction learning and exposure outcome, but not (necessarily) that enhanced extinction is related to better exposure outcome (Geller et al., 2019; Hahn et al., 2015). Finally, two studies did not find evidence for an association between extinction learning and exposure treatment response (Raeder et al., 2020; Wannemueller et al., 2018).

Extinction recall

Forcadell et al. (2017) examined, additionally to extinction, whether individual differences in extinction recall in the laboratory are associated with the outcome of an exposure treatment analog. One day after extinction (before treatment took place), extinction recall was tested in the extinction context (B) by switching the lamp to the CS+ and CS- color without occurrence of the US. When analyzing the data of the entire sample, extinction recall (i.e., CS+/CS- differentiation during the first two recall trials) was not associated with exposure outcome. In a selection of participants that showed acquisition and extinction learning, participants with less extinction recall in FPS (i.e., CS+/CS- differentiation during the first two extinction recall trials) showed larger percentages of fear reduction in FPS and pre-post reductions in FPS. Notably, these results are opposite to what was predicted, i.e., that better extinction recall would be associated with better exposure outcome.

Similar to Forcadell et al. (2017), Raeder et al. (2020) included an extinction recall phase in their fear conditioning task. The extinction recall phase tested for contextual renewal or a return in fear responding due to a shift in context after extinction. Extinction recall was tested after a 10-minute interval (after extinction) by presenting the CS+ and CS- without the

US in both context A and B. Non-completers attached a more negative valence to the CSs and reported higher CS-US contingency ratings in the extinction recall phase compared to responders. This effect remained after controlling for differences at the end of extinction. However, contrary to what was hypothesized, no context-dependent increase in fear responding was observed between the end of extinction and the extinction recall phase. Instead, during extinction recall, an overall further decrease in CS-US contingency ratings was observed. This suggests that impairments in the further extinction during extinction recall are associated with a failure to accomplish exposure within a predetermined time.

In conclusion, the available evidence on predicting exposure treatment response by extinction recall is limited and not consistent. Forcadell et al. (2017) (only in a subgroup of their sample) found, opposite to what was hypothesized, that less extinction recall is associated with better exposure outcome. Raeder et al. (2020) found that less fear reduction during extinction recall is associated with non-completion of exposure treatment.

Generalization

No results were found with regard to predicting exposure-based treatment response by generalization in fear conditioning.

Acquisition

Waters and Pine (2016), in addition to extinction, examined whether fear acquisition predicted CBT response. Responders, non-responders and non-anxious controls did not differ in the acquisition of differential SC. With regard to the ratings, non-responders rated both CSs as more unpleasant after acquisition compared to before acquisition, whereas valence ratings did not change across phases in treatment responders and the non-anxious control group. No group differences were found in the arousal ratings from pre-to post-acquisition. No significant correlations were found between continuous symptom change and the rating data after acquisition. The authors conclude that the acquisition pattern of treatment responders

resembled the pattern of the non-anxious controls: both groups did not acquire conditioned negative evaluations, whereas non-responders did acquire negative valence after acquisition.

Geller et al. (2009) found that responders and non-responders to CBT show similar acquisition of fear in SC. Raeder et al. (2020) found that completers and non-completers did not differ in fear acquisition. Similarly, Ball et al. (2016) found that valence and arousal ratings after acquisition do not predict exposure outcome.

Wannemueller and colleagues (2018) also investigated whether fear acquisition patterns are correlated with the outcome of their standardized exposure-based treatment. Based on participants' responses to the dichotomous (yes/no) assessment of contingency at early acquisition, three learning patterns were distinguished. Participants who neither expected the US after the CS+ nor CS- or expected the US to follow after the CS- but not after the CS+ were identified as *poor learners*. Participants who expected the US after the CS+ but not after the CS- were identified as *accurate learners*. Participants who expected the US to follow both the CS+ and CS- were identified as *threat-biased learners*. Participants in the threat-biased learning group showed larger fear reductions from pre- to post-treatment compared to accurate learners. There were no differences between the learning groups in fear reduction at follow-up.

Finally, Hahn et al. (2015), using a multivariate pattern classification approach, found that neural activation during acquisition enabled 74 percent correct responder versus non-responder classifications.

In conclusion, with regard to evidence on whether fear acquisition predicts exposure-based treatment response most of the studies do not find an association between acquisition and exposure treatment response (Ball et al., 2016: Geller et al., 2019; Raeder et al., 2020; Waters & Pine, 2016).

Avoidance

No studies on predicting exposure-based treatment response by avoidance learning in lab-based fear conditioning were found.

Conclusion and Discussion

One of the goals of laboratory-based fear conditioning research is to improve our understanding of the onset and treatment of clinical anxiety. A prerequisite for this, however, is that fear conditioning is a valid model of which the results can be generalized to clinical anxiety. Prospective research provides one important way to empirically evaluate the external validity of the fear conditioning model. It concerns testing whether individual differences in performance in fear conditioning in the laboratory can predict individual differences in (clinical) anxiety. In this paper, we provided an overview of available prospective studies using laboratory-based fear conditioning to predict (1) anxiety levels and (2) exposure-based treatment outcome. We focused on extinction, generalization, acquisition and avoidance in lab-based fear conditioning as predictors. Our literature search resulted in five studies on the predictive value of lab-based fear conditioning for (clinical) anxiety levels and seven studies investigating whether fear conditioning can predict exposure-based treatment outcome.

With respect to predicting anxiety symptoms by lab-based *extinction*, three studies were found. The results of the reviewed studies suggest that reduced extinction learning is predictive of higher anxiety levels (Guthrie & Bryant, 2006; Lommen et al., 2014; Orr et al., 2012). Five prospective studies were found on lab-based *acquisition* as a predictor for anxiety levels, resulting in mixed evidence. Some of these studies showed an association between acquisition and later anxiety, whereas other studies did not (Guthrie & Bryant, 2006 in SCR; Sijbrandij et al., 2013). Moreover, amongst the studies that do find a significant correlation, some studies show that increased US-expectancy ratings to the CS- and impaired CS+/CS- discrimination predict higher anxiety (Lenaert et al., 2014), whereas other studies report that higher CS+ responding or larger CS+/CS- discrimination is related to higher anxiety levels

(see results on EMG responding in Guthrie & Bryant, 2006; Orr et al. 2012). Only one study was found on predicting subclinical anxiety levels by lab-based *generalization*. Generalization of US-expectancy ratings to stimuli that resemble the CS- was found to be predictive for higher anxiety levels (Lenaert et al., 2014). Notably, this is the only study that used aversive pictures as US. Also, subclinical anxiety levels were measured in a student sample instead of clinical anxiety in high-risk samples as is the case in most other studies on this topic. No studies were found using lab-based *avoidance* to predict anxiety symptoms.

With regard to predicting exposure-based treatment outcome, seven studies focus on lab-based *extinction* learning. Again, some studies do find an association between extinction and exposure-based treatment outcome (Ball et al., 2016; Forcadell et al., 2017; Geller et al., 2019; Hahn et al., 2015; Waters & Pine, 2016), whereas other studies do not (Raeder et al., 2020; Wannemueller et al., 2018). Moreover, amongst the studies showing an association, some provide evidence that *better* extinction learning (i.e., larger decline in CS+/CS- differentiation during extinction, smaller differences between the CS+ and CS- in valence ratings after extinction) predicts better treatment outcome (Ball et al., 2016; Forcadell et al., 2017; Waters & Pine, 2016), whereas other studies find that treatment responders showed larger CS+/CS- differentiation throughout extinction (Geller et al., 2019) or that brain activation patterns during extinction predicts treatment response (Hahn et al., 2015). Two studies were found on predicting exposure treatment outcome by lab-based *extinction recall*. Opposite to what was predicted, one study showed, in a subset of their sample, that less extinction recall is associated with better treatment outcome (Forcadell et al., 2017). The other study failed to find contextual renewal, but showed that less fear reduction during the extinction recall phase is associated with not completing treatment. No studies were found on predicting exposure treatment outcome based on lab-based *generalization* or *avoidance*. Six studies report results on predicting exposure-based treatment outcome by lab-based

acquisition. The majority of studies failed to find an association (Ball et al., 2016: Geller et al., 2019; Raeder et al., 2020; Waters & Pine, 2016). One study found that participants who expected the US to follow both the CS+ and CS- at the end of acquisition respond better to treatment (Wannemueller et al., 2018). Another study showed that brain activation during acquisition could reliably predict who will respond to treatment (Hahn et al., 2015).

It is difficult to draw solid conclusions from the published prospective studies for several reasons. First, it is remarkable that, despite the importance of this research topic, a total of only twelve prospective studies was found. In addition, some learning processes are understudied. We found only one prospective study on generalization (i.e., Lenaert et al., 2014) and no prospective studies on avoidance. The time-consuming nature of prospective research might (partially) explain the limited amount of available studies. However, it is also possible that not all of the conducted work on this topic has been published. Remarkably, all published studies contain at least some significant results and no studies with only null findings were published. It is possible that at least some conducted work on this topic – in particular resulting in null findings – did not find its way to publication. A search for unpublished data could provide more certainty about the existence and extent of unpublished null findings, which might be useful for future systematic reviews and meta-analyses on this topic.

Second, most of the studies rely on small sample sizes. Some studies initially included larger sample sizes, but sample sizes in these studies were significantly reduced after participant drop-out (e.g., Geller et al., 2019) or exclusion due to acquisition or extinction criteria (e.g., Guthrie & Bryant, 2006; Lommen et al., 2013). Particularly for correlational designs, this could result in underpowered studies, increasing the risk of false conclusions. Studies with low statistical power have both a reduced chance of detecting a true effect (type II error) as well as a reduced likelihood that a significant result reflects a true effect (type I

error) (Button et al., 2013). Additionally, given that after an aversive (conditioning) experience in real life typically only a minority of individuals develops an anxiety disorder, large sample sizes are required to examine whether fear conditioning predicts the onset of an anxiety disorder. Some studies included at risk populations (e.g., soldiers, firemen) to increase the likelihood of the onset of an anxiety disorder. However, even in these studies, PTSD diagnosis could not be used as an outcome measure because of the limited number of participants that actually developed PTSD (e.g., Guthrie & Bryant, 2006; Lommen et al., 2013; Sijbrandij et al., 2013).

Third, it is difficult to draw solid overarching conclusions because direct comparison between studies is not straightforward. First, acquisition and extinction learning are operationalized in very heterogeneous ways across studies. For instance, Lommen et al. (2013) operationalized extinction learning as the reduction in CS+ responding during early extinction (i.e., from trial 1 to 4), whereas Orr et al. (2012) and Guthrie and Bryant (2006) looked at CS+/CS- discrimination throughout the entire extinction phase. It is not always clear why certain operationalizations are chosen. Second, studies differ in which measures of conditioning (e.g., expectancy ratings, SC, FPS, EMG, arousal ratings, valence ratings) are found to be predictive for anxiety symptoms or treatment outcome. Most studies include multiple measures during fear conditioning and significant effects are often observed in only one or a subset of the fear measures. The same holds for the outcome measures that have been included to assess (clinical) anxiety levels and the operationalization of treatment outcome. Typically, multiple outcome measures are included (e.g., self-report symptom measures, diagnostic interview, physiological data) and significant associations are obtained for only one or a subset of outcome measures. Conclusions, however, often focus on the significant results and disregard the non-significant results in the other measures.

Different operationalizations of learning processes, the inclusion of multiple conditioning measures and multiple outcome measures give rise to an abundance of possible combinations and hypotheses that can be tested (Lonsdorf et al., 2019). Most studies, however, do not specify whether decisions about the primary hypotheses are made beforehand or post-hoc (after inspection of the data). If primary hypotheses are not selected beforehand, this increases the researcher's degree of freedom in focusing on and selecting the significant results from a larger set of analyses and disregarding null findings. Moreover, studies do not always apply appropriate statistical methods to account for the inclusion of multiple measures, increasing the risk of false positive correlations. For future research on this topic, we encourage preregistration of the primary and main hypotheses, transparency about all conducted analyses and adequate correction for multiple tests.

Given the observation that the effects of the included studies might be sensitive to and depend on (1) the operationalization of learning processes, (2) the measure of fear during conditioning, (3) the outcome measure and (4) the exclusion of participants, it can be questioned how robust these effects are. Performing a multiverse analysis can increase transparency and provide an idea of the robustness of the effect (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). In a multiverse analysis, analyses across a set of reasonable processing choices and combinations are performed. A distribution of the resulting *p*-values or effects sizes of all conducted analyses is reported to provide an overview of the effect. As discussed earlier, the choice of a particular way of processing the data (e.g., excluding participants, operationalization of the learning effect) often relies on arbitrary grounds. A multiverse analysis can provide insights in how much the conclusions change because of such arbitrary choices.

Prospective research on the predictive value of fear conditioning for the onset and treatment of (clinical) anxiety can serve different goals. First, if performance in the laboratory

and anxiety symptoms in real life are moderated by the individual in a similar way, this provides empirical evidence for the external validity of the fear conditioning model. Second, from a theoretical standpoint, this type of research can increase our understanding of the candidate processes involved in the pathogenesis and treatment of clinical anxiety. Third, from a practical standpoint, this research can provide empirical support for fear conditioning as a prognostic test or marker to identify individuals at risk for developing an anxiety disorder. This could be particularly useful in high-risk groups such as soldiers, fire fighters, policemen, etc. In addition, if individual differences in fear conditioning reliably predict treatment outcome, fear conditioning can be used as a prognostic marker in guiding treatment selection. Given that approximately 45 percent of patients suffering from an anxiety disorder do not fully benefit from CBT, it seems worthwhile to identify beforehand which patients are likely to benefit from CBT and which patients might benefit more from another intervention (Loerinc et al., 2015; Rapee, Schniering, & Hudson, 2009). At the same time, we should perhaps not overvalue the potential usefulness of fear conditioning as a marker. An ideal marker should be able to predict the future occurrence of a disorder at an individual level, whereas the predictive effects of fear conditioning remain small and inconsistent even at a group level. In addition, to be feasible, a marker should be easy to measure, whereas fear conditioning – at least if physiological measures are included – requires specific equipment and can be time-consuming.

In this type of prospective research – and in fear conditioning research more general – a variety of stimuli is used. Typically, it is assumed that individual differences do not depend on the exact stimulus material that is used. However, it might be the case that individual differences interact with the stimulus material (e.g., spiders in spider phobics). Moreover, the nature of aversive events experienced outside the laboratory might differ greatly between participants. At risk samples such as firemen or soldiers have been used so that the aversive

events are more or less the same within the sample. However, even in these studies there might be differences in the nature of the traumatic experience, for instance whether a soldier was 10m away or 100m away when a grenade was dropped. These differences are typically not taken into account in prospective studies, but might strongly affect the results (similar to the influence of US intensity in fear conditioning studies; De Houwer & Hughes, 2020).

In conclusion, there is a need for more prospective research with sufficiently large sample sizes, clearly preregistered hypotheses and transparency with regard to the conducted analyses. In particular, more research is needed on the prospective value of generalization and avoidance. Additionally, future research can focus on whether combinations of learning processes have better predictive power. For instance, someone who overgeneralizes and shows slow extinction might be more vulnerable to develop an anxiety disorder compared to someone who shows slow extinction but does not overgeneralize. Importantly, also high-quality studies with null findings should find their way to publication to provide us with a complete overview of existing research and a more reliable estimation of the true effects.

References

*Ball, T. M., Knapp, S. E., Paulus, M. P., & Stein, M. B. (2017). Brain activation during fear extinction predicts exposure success. *Depression and Anxiety, 34*, 257-266. doi:10.1002/da.22583

Beckers, T., Krypotos, A. M., Boddez, Y., Effting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biological Psychology, 92*, 90-96. doi:10.1016/j.biopsycho.2011.12.015

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermine the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376. doi:10.1038/nrn3475

Craske, M. G., Hermans, D., & Vansteenwegen, D. (Eds.). (2006). *Fear and learning: From basic processes to clinical implications*. American Psychological Association. doi:10.1037/11474-000

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. doi:10.1016/j.brat.2020.103558

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 16*(2), 237-242. doi:10.1037/0097-7403.26.2.237

De Houwer, J., & Hughes, S. (2020). *The psychology of learning: An introduction from a functional-cognitive perspective.* Cambridge, MA: MIT Press.

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., …, & Baas, J. M. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety, 32*, 239-253. doi:10.1002/da.22353

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear generalization in humans: Systematic review and implications for anxiety disorder research. *Behavior Therapy, 46*, 561-582. doi:10.1016/j.beth.2014.10.001

Dyson, R., & Renk, K. (2006). Freshmen adaptation to university life: Depressive symptoms, stress, and coping. *Journal of Clinical Psychology, 62*, 1231-1244. doi:10.1002/jclp.20295

*Forcadell, E., Torrents-Rodas, D., Vervliet, B., Leiva, D., Tortella-Feliu, M., & Fullana, M. A. (2017). Does fear extinction in the laboratory predict outcomes of exposure therapy? A treatment analog study. *International Journal of Psychophysiology, 121*, 63-71. doi:10.1016/j.ijpsycho.2017.09.001

*Geller, D. A., McGuire, J. F., Orr, S. P., Small, B. J., Murphy, T. K., Trainor, K. …, & Storch, E. A. (2019). Fear extinction learning as a predictor of response to cognitive behavioral therapy for pediatric obsessive compulsive disorder. *Journal of Anxiety Disorders, 64*, 1-8. doi:10.1016/j.janxdis.2019.02.005

Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Aspergis-Schoute, A. M., …, & Robbins, T. W. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biological Psychiatry, 75*, 631-638. doi:10.1016/j.biopsych.2013.02.002

*Guthrie, R. M., & Bryant, R. A. (2006). Extinction learning before trauma and subsequent posttraumatic stress. *Psychosomatic Medicine, 68*, 307-311. doi:10.1097/01.psy.0000208629.67653.cc

*Hahn, T., Kircher, T., Straube, B., Wittchen, H. U., Konrad, C., Ströhle, A., …, & Lueken, U. (2015). Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. *JAMA Psychiatry, 72*, 68-74. doi:10.1001/jamapsychiatry.2014.1741

Hermans, D., Boddez, Y., & Vervliet, B. (2019). The experimental study of mental health

problems – Building on Watson & Rayner (1920). In G. Davey (Ed.), *Classic studies in clinical psychology* (27-40). California: Sage Publishing.

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biological Psychiatry, 60*, 361-368. doi:10.1016/j.biopsych.2005.10.006

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*, 337–343. doi:10.1001/archpsyc.1997.01830160065009

Krypotos, A.-M., Vervliet, B., & Engelhard, I. M. (2018). The validity of human avoidance paradigms. *Behaviour Research and Therapy, 111,* 99-105. doi:10.1016/j.brat.2018.10.011 LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*, 155-184. doi:10.1146/annurev.neuro.23.1.155

*Lenaert, B., Boddez, Y., Griffith, J. W., Vervliet, B., Schruers, K., & Hermans, D. (2014). Aversive learning and generalization predict subclinical levels of anxiety: A six-month longitudinal study. *Journal of Anxiety Disorders, 28*, 747-753. doi:10.1016/j.anxdis.2014.09.006

Lissek, S., Biggs, A. L., Rabin, S. J., Cornwell, B. R., Alvarez, R. P., Pine, D. S., & Grillon, C. (2008). Generalization of conditioned fear-potentiated startle in humans. *Behaviour Research and Therapy, 46*, 678-687. doi:10.1016/j.brat.2008.02.005

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. *Behaviour Research and Therapy, 43*, 1391-1424.

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C.

(2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *American Journal of Psychiatry, 167*, 47-55. doi:10.1176/appi.ajp.2009.09030410

Loerinc, A. G., Meuret, A. E., Twohig, M. P., Rosenfield, D., Bluett, E. J., & Craske, M. G. (2015). Response rated for CBT for anxiety disorders: Need for standardized criteria. *Clinical Psychology Review, 42*, 72-82. doi:10.1016/j.cpr.2015.08.004

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

Lonsdorf, T. B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., …, & Merz, C. J. (2019). Navigating the garden of forking paths for data exclusions in fear conditioning research. *eLife, 8*, e52465. doi:10.7554/eLife.52465

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., … & 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 and Biobehavioral Reviews, 77*, 247-285. doi:10.1016/j.neurobiorev.2017.02.026

Lovibond, P. F., Mitchell, C. J., Minard, e., Brady, A., & Menzies, R. G. (2009). Safety behaviors preserve threat beliefs: Protection from extinction of human fear conditioning by an avoidance response. *Behaviour Research and Therapy, 47*, 716-720. doi:10.1016/j.brat.2009.04.013

McNally, R. J. (2007). Mechanisms of exposure therapy: How neuroscience can improve psychological treatments for anxiety disorders. *Clinical Psychology Review, 27*, 750-759. doi:10.1016/j.cpr.2007.01.003

Milad, M. R., Orr, S. P., Pitman, R. K., & Rauch, S. L. (2005). Context modulation of

memory for fear extinction in humans. *Psychophysiology, 42*, 456-464. doi:10.111/j.1469-8986.2005.00302.x

Milad, M. R., & Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature, 420*, 70-74. doi:10.1038/nature01138

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*, 10-26. doi:10.1037/0003-066X.61.1.10

*Orr, S. P., Lasko, N. B., Macklin, M. L., Pineles, S. L., Chang, Y., & Pitman, R. K. (2012). Predicting post-trauma stress symptoms from pre-trauma psychophysiologic reactivity, personality traits and measures of psychopathology. *Biology of Mood & Anxiety Disorders, 2*, 8. doi:10.1186/2045-5380-2-8

Öst, L. G., Havnen, A., Hansen, B., & Kvale, G. (2015). Cognitive behavioral treatments of obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993-2014. *Clinical Psychology Review, 40*, 156-169. doi:10.1016/j.cpr.2015.06.003

Pöhlchen, D., Leuchs, L., Binder, F. P., Blaskovich, B., Nantawisarakul, T., Topalidis, P., …, & Spoormaker, V. I. (2020). No robust differences in fear conditioning between patients with fear-related disorders and healthy controls. *Behaviour Research and Therapy, 129*, 103610. doi:10.1016/j.brat.2020.103610

*Raeder, F., Merz, C. J., Margraf, J., & Zlomuzica, A. (2020). The association between fear extinction, the ability to accomplish exposure and exposure therapy in specific phobia. *Scientific Reports, 10*, 4288. doi:10.1038/s41598-020-61004-3

Rapee, R. M., Schniering, C. a., & Hudson, J. L. (2009). Anxiety disorders during childhood and adolescence: Origins and treatment. *Annual Review of Clinical Psychology, 5*, 311-341. doi:10.1146/annurev.clinpsy.032408.153628

Salters-Pedneault, K., Tull, M. T., & Roemer, L. (2004). The role of avoidance of emotional material in the anxiety disorders. *Applied Preventive Psychology, 11*, 95-114. doi:10.1016/j.appsy.2004.09.001

Scheveneels, S., Boddez, Y., & Hermans, D. (2019). Learning mechanisms in fear and anxiety: It is still not what you think it is. In B. Olatunji (Ed.), *The Cambridge Handbook of Anxiety and Related Disorders* (pp. 13-40). Cambridge University Press. doi:10.1017/9781108140416.002

Scheveneels, S., Boddez, Y., Vervliet, B., & Hermans, D. (2016). The validity of Laboratory-based treatment research: Bridging the gap between fear extinction and exposure treatment. *Behaviour Research and Therapy, 86*, 87-94. doi:10.1016/j.brat.2016.08.015

*Sijbrandij, M., Engelhard, I. M., Lommen, M. J. J., Leer, A., & Baas, J. M .P. (2013). Impaired fear inhibition learning predicts the persistence of symptoms of posttraumatic stress disorder (PTSD). *Journal of Psychiatric Research, 47*, 1991-1997. doi:10.1016/j.psychires.2013.09.008

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*, 702-7012. doi:10.1177/1745691919958637

Vervliet, B., & Boddez, Y. (2020). Memories of 100 years of human fear conditioning research and expectations for its future. *Behaviour Research and Therapy, 135*, 103732. doi: 10.1016/j.brat.2020.103732

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

Vervliet, B., & Raes, F. (2013). Criteria of validity in experimental psychopathology:

Application to models of anxiety and depression. *Psychological Medicine, 43*, 2241-2244. doi:10.1017/S0033291712002267

Vervliet, B., Vansteenwegen, D., & Eelen, P. (2006). Generalization gradients for acquisition and extinction in human contingency learning. *Experimental Psychology, 53*, 132-142. doi:10.1027/1618-3169.53.2.132

*Wannemueller, A., Moser, D., Kumsta, R., Jöhren, H.-P., Adolph, D., & Margraf, J. (2018). Mechanisms, genes and treatment: Experimental fear conditioning, the serotonin transporter gene, and the outcome of a highly standardized exposure-based fear treatment. *Behaviour Research and Therapy, 107*, 117-126. doi:10.1016/j.brat.2018.06.003

*Waters, A. M., & Pine, D. S. (2016). Evaluating differences in Pavlovian fear acquisition and extinction as predictors of outcome from cognitive behavioural therapy for anxious children. *The Journal of Child Psychology and Psychiatry, 57*, 869-876. doi:10.1111.jcpp.12522

Watson, J. B., & Rayner, R; (1920). Conditioned emotional reactions. *Journal of Experimental Psychology, 3*, 1-14.

Whitfield-Gabrieli, S., Gosh, S. S., Nieto-Castanon, A., Saygin, Z., Doehrmann, O., Chai, X. J., …, & Gabrieli, J. D. (2016). Brain connectomics predict response to treatment in social anxiety disorder. *Molecular Psychiatry, 21*, 680-685. doi:10.1038/mp.2015.109

Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. *Clinical Psychology Review, 28*, 1021-1037. doi:10.1016/j.cpr.2008.02.007