Treatment-like use of discrimination training to reduce generalization of conditioned fear

Behandlungsähnlicher Einsatz eines Diskriminationstrainings zur Verringerung von Generalisierung konditionierter Furcht

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Neuroscience

submitted by

Katharina Hutterer, née Herzog

from

Bayreuth

Würzburg 2022

Submitted on:

Members of the Thesis Committee

Chairperson:	Prof. Dr. Matthias Gamer
Primary Supervisor:	Prof. Dr. Paul Pauli
Supervisor (Second):	Prof. Dr. Jürgen Deckert
Supervisor (Third):	PD Dr. Robert Blum
Supervisor (Fourth):	Prof. Dr. Marta Andreatta

Acknowledgements

Diese Doktorarbeit ist im Rahmen des DFG geförderten Sonderforschungsbereichs TRR 58 Projekt Z02 entstanden. Dafür danke ich meinem Promotionskomitee, Prof. Dr. Paul Pauli, Prof. Dr. Jürgen Deckert, PD Dr. Robert Blum und Prof. Dr. Marta Andreatta. Sie haben mir die Mitarbeit an diesem standortübergreifenden, interdisziplinären Projekt ermöglicht. Dabei konnte ich mich mithilfe Ihrer exzellenten wissenschaftlichen Beratung und unseres inspirierenden inhaltlichen Austauschs in das Thema Furcht, Angst und Angsterkrankungen vertiefen. Gleichzeitig erlaubte mir das Projekt auch den Austausch mit vielen anderen (Jung-)WissenschaftlerInnen auf diesem Forschungsgebiet. Dies ist nur ein Beispiel, wie Sie meine wissenschaftliche Entwicklung auch projektunabhängig gefördert haben. Vielen Dank für diese Unterstützung und Förderung.

Danke auch an meine studentischen Hilfskräfte und Studierenden, ohne deren Hilfe die Rekrutierung und Untersuchung der gewaltigen Versuchspersonenzahl nicht möglich gewesen wäre.

Großer Dank gilt außerdem meinen Kolleginnen und Kollegen am Lehrstuhl 1 für Psychologie. Eure wissenschaftliche Neugier und Skepsis sind ansteckend und bereichernd. Unser gegenseitiges Interesse für unsere Arbeit und die resultierende gegenseitige Unterstützung sind beispielhaft und waren für mich ein Riesengewinn.

Danke an all die wundervollen Menschen innerhalb und außerhalb meiner Arbeitsumgebung, die mich auf dem Weg hierher begleitet haben, die mich auch in stressigen Zeiten ertragen oder geduldig auf mich verzichtet haben, deren Freundschaft und Liebe mir so viel Kraft und Durchhaltevermögen geschenkt hat.

Table of contents

List of figures

List of tables

Summary

Anxiety patients overgeneralize fear, also because of an inability to perceptually discriminate threat and safety signals. Therefore, some studies have developed discrimination training that successfully reduced the occurrence of fear generalization. The present work is the first to take a treatment-like approach by using discrimination training after generalization has occurred. Therefore, two studies were conducted with healthy participants using the same fear conditioning and generalization paradigm, with two faces as conditioned stimuli (CSs), and four facial morphs between CSs as generalization stimuli (GSs). Only one face (CS+) was followed by a loud scream (unconditioned stimulus, US). In Study 1, participants underwent either fear-relevant (discriminating faces) or fear-irrelevant discrimination training (discriminating width of lines) or a non-discriminative control training between the two generalization tests, each with or without feedback (n = 20 each). Generalization of US expectancy was reduced more effectively by fear-relevant compared to fear-irrelevant discrimination training. However, neither discrimination training was more effective than non-discriminative control training. Moreover, feedback reduced generalization of US expectancy only in discrimination training. Study 2 was designed to replicate the effects of the discrimination-training conditions in a large sample (N = 244) and examine their benefits in individuals at risk for anxiety disorders. Again, feedback reduced fear generalization particularly well for US expectancy. Fear relevance was not confirmed to be particularly fear-reducing in healthy participants, but may enhance training effects in individuals at risk of anxiety disorder. In summary, this work provides evidence that existing fear generalization can be reduced by discrimination training, likely involving several (higher-level) processes besides perceptual discrimination (e.g., motivational mechanisms in feedback conditions). Its use may be promising as part of individualized therapy for patients with difficulty discriminating similar stimuli.

Zusammenfassung

Angstpatienten übergeneralisieren Furcht, unter anderem weil sie nicht in der Lage sind, Bedrohungs- und Sicherheitsreize zu unterscheiden. Daher wurde in einigen Studien ein Diskriminationstraining entwickelt, das das Auftreten von Furchtgeneralisierung erfolgreich reduzierte. Die vorliegende Arbeit ist die erste, die einen behandlungsähnlichen Ansatz verfolgt, indem sie Diskriminationstraining einsetzt, nachdem die Generalisierung stattgefunden hat. Zu diesem Zweck wurden zwei Studien mit gesunden Teilnehmern durchgeführt, die dasselbe Paradigma zur Furchtkonditionierung und -generalisierung verwendeten, mit zwei Gesichtern als konditionierte Stimuli (CSs) und vier Gesichtsmorphen zwischen den CS als Generalisierungsstimuli (GSs). Nur auf ein Gesicht (CS+) folgte ein lauter Schrei (unkonditionierter Stimulus, US). In Studie 1 durchliefen die Teilnehmer zwischen den beiden Generalisierungstests entweder ein furchtrelevantes (Unterscheidung von Gesichtern) oder ein furchtirrelevantes Diskriminationstraining (Unterscheidung der Breite von Linien) oder ein non-diskriminatives Kontrolltraining, jeweils mit oder ohne Feedback (jeweils $n = 20$). Die Generalisierung der US-Erwartung wurde durch furchtrelevante im Vergleich zu furchtirrelevanten Diskriminationstrainings effektiver reduziert. Keines der beiden Diskriminationstrainings war jedoch effektiver als ein non-diskriminatives Kontrolltraining. Darüber hinaus verringerte das Feedback die Generalisierung der US-Erwartung nur im Diskriminationstraining. Studie 2 sollte die Effekte der Diskriminationstrainingsbedingungen in einer großen Stichprobe ($N = 244$) replizieren und ihre Effekte bei Individuen mit einem Risiko für Angststörungen untersuchen. Auch hier reduzierte das Feedback die Furchtgeneralisierung besonders gut für die US-Erwartung. Die Furchtrelevanz erwies sich bei gesunden Teilnehmern nicht als besonders furchtreduzierend, könnte aber die Trainingseffekte bei Personen

mit einem Risiko einer Angststörung verstärken. Zusammenfassend lässt sich sagen, dass diese Arbeit Hinweise dafür liefert, dass bestehende Furchtgeneralisierung durch ein Diskriminationstraining reduziert werden kann, wobei wahrscheinlich mehrere Prozesse (höherer Ordnung) neben der perzeptuellen Diskrimination beteiligt sind (z. B. motivationale Mechanismen in den Feedback Bedingungen). Die Anwendung des Diskriminationstrainings als Teil einer individualisierten Therapie für Patienten mit Schwierigkeiten bei der Unterscheidung ähnlicher Stimuli könnte vielversprechend sein.

1 Introduction

"Be fearful of mediocrity." – Jonathan Ellery

Usually, mediocrity is seen as not worth striving for, as boring, as falling short of the possibilities. Anyone who wants to do better should beware of it, fear it. Interesting choice of words. Where precisely on the subject of fear and anxiety, a healthy mediocrity is the most desirable state.

Fear is an adaptive emotion (Davis, Walker, Miles, & Grillon, 2010; Lang & Bradley, 2010), without which we would be less able to respond to threats and survival would be less assured (Hamm, 2020; Lang and Bradley, 2010). Thanks to the mechanism of generalization, prior fear learning can be extended when confronted with other potentially threatening stimuli, allowing to decide more quickly how to react (Lissek et al., 2008; Onat & Buchel, 2015; Schiele et al., 2016).

However, when becoming excessive, i.e., if fear is overgeneralized, it becomes maladaptive, leading to pronounced avoidance behavior with a significant reduction in quality of life. At worst, this can lead to the development of an anxiety disorder (American Psychiatric Association, 2013; Mineka & Zinbarg, 2006), which are the most common mental disorders in Europe and the United States (Bandelow & Michaelis, 2015).

Therefore, it is of high clinical relevance to further investigate the development of excessive fear generalization and to find out how existing fear generalization can be reduced again. The present dissertation aims to contribute to this by investigating whether generalization of conditioned fear that has already occurred can be reduced by a discrimination training.

1.1 Emotional responses

Emotions have several functions, e.g., they recruit response systems, motivate and coordinate cognition and action, provide information or meaning to facilitate decision-making, influence social relationships (Izard, 2010). Different emotions have different functions, but all functions can contribute to physical and mental health (Izard, 2010). Ekman and colleagues have identified six basic emotions that are expressed with distinctive facial expressions across cultures (Ekman & Friesen, 1971, 1975; for a review on basic emotions also see Tracy & Randles, 2011). These basic emotions include surprise, fear, disgust, anger, happiness and sadness.

Of course, emotions are expressed not only by a pronounced facial expression, but also by reactions at the physiological, verbal (i.e., subjective), and behavioral levels (Bradley & Lang, 2000; Lang, 1968). One possibility to conduct the physiological level of emotional response is the recording of electro-dermal activity (EDA). The EDA reacts to emotional and/or salient stimuli by a phasic increase, the so-called skin conductance response (SCR; Boucsein et al., 2012; Lonsdorf et al., 2017). For example, Lang, Greenwald, Bradley, and Hamm (1993) showed that skin conductance responses to pictures (international affective picture system, IAPS, Lang, Bradley, & Cuthbert, 2008) increase the more arousing a picture was rated. Banks, Bellerose, Douglas, and Jones-Gotman (2012) examined SCR amplitudes to emotionally expressive faces while participants completed a forced-choice task in which the emotionality of the facial images was irrelevant, finding stronger responses to sad and happy faces compared to neutral faces from the left hand. Another study presented emotional faces with varying degrees of anxiety and found a linear relationship between participants' skin conductance responses and increasing intensity of negative emotionality (Fusar-Poli, Landi, & O'Connor, 2009). Alternatively, one can record the startle response, which is a

defense eye-blink reflex and is potentiated under threatening conditions (Andreatta, Muhlberger, Yarali, Gerber, & Pauli, 2010; Lonsdorf et al., 2017). The most obvious measure regarding the verbal level are ratings (Lonsdorf et al., 2017). Hereby, the verbal level can be further divided in cognitive and affective components (Kleinginna & Kleinginna, 1981; Lonsdorf et al., 2017). The cognitive component encompasses appraisal and labeling processes, and for example can be measured by US-expectancy ratings, while the affective component indicates the degree of pleasantness and activation (Kleinginna & Kleinginna, 1981; Lonsdorf et al., 2017). According to the Circumplex Model (Barrett & Russell, 2016; Posner, Russell, & Peterson, 2005; Russell & Barrett, 1999), affective feelings involved in emotions can be described by these two bipolar dimensions, arranged orthogonally. The pleasantness dimension summarizes hedonic tone, i.e., how pleasant or unpleasant an affect is evaluated. The dimension activation refers to the arousal or sense of energy while experiencing an affect (Barrett & Russell, 2016). Importantly, prototypical emotional responses can be localized in certain regions of the circumplex (Russell & Barrett, 1999). Consequently, it is appropriate to measure the affective component of emotional responses with ratings of valence and arousal (Lonsdorf et al., 2017).

Fear, for instance, is characterized by a combination of strong activation and negative valence (Posner et al., 2005; Russell & Barrett, 1999). The high activation level contributes to physiological mobilization and readiness for action, which is an important function of fear (Hamm, 2020; Izard, 2010; Lang and Bradley, 2010), because fear evolves in reaction to an imminent threat and real danger. Thereby, fear is evoked rapidly, but also vanishes rapidly as soon as the threat is removed. Consequently, fear is an adaptive, temporal state (Davis et al., 2010; Lang & Bradley, 2010) that allows us to predict or respond (e.g., with fight or flight) to an imminent or specific threat and adapt our behavior to past, present, and future threats (Dymond,

Dunsmoor, Vervliet, Roche, & Hermans, 2015). In contrast, a state of apprehension due to less specific, less predictable or more distant threats is termed anxiety. It is characterized by increased arousal and vigilance in preparation to a future potential threat and is accompanied by cautious or avoiding behavior (American Psychiatric Association, 2013; Davis et al., 2010). Thus, both fear and anxiety aim to achieve action disposition in order to cope with threat and ensure survival (Hamm, 2020; Lang and Bradley, 2010). However, if fear or anxiety are experienced excessively, they become maladaptive, as they lead to pronounced avoidance behavior along with substantial restriction in quality of life. At worst, this can result in the development of an anxiety disorder (American Psychiatric Association, 2013; Mineka & Zinbarg, 2006).

According to several surveys of representative populations in the United States and Europe, anxiety disorders are the most common mental illnesses, with a 12-month prevalence ranging from 8.3% to 21.3% (Bandelow & Michaelis, 2015). Anxiety disorders are more frequently observed in females than in males (ratio 2:1; American Psychiatric Association, 2013; Bandelow & Michaelis, 2015). The psychopathology of the respective anxiety disorders is described in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013). According to the manual, some anxiety disorders are characterized by an intense fear of specific objects or situations. This applies, for example, to specific phobia (e.g., fear of spiders or height), agoraphobia (which is fear of situations where escape might be difficult or help would not be available) or social anxiety disorder (SAD, characterized by fear of social situations or interactions). Core symptom of a panic disorder is the repeated occurrence of unexpected panic attacks, i.e., sudden occurring states of overwhelming fear and intense apprehension, which are accompanied by severe physical and cognitive symptoms. Affected persons are permanently concerned to have another panic attack or show inappropriate behavior to avoid these. Patients with generalized anxiety disorder

(GAD) are suffering from sustained and excessive anxiety and worries that are hard to control and affect various areas of daily life.

Current etiological models point out that many anxiety disorders arise from the interaction of vulnerability factors (i.e., diatheses) and stressors (e.g., Kendler, Myers, & Prescott, 2002; Mineka & Zinbarg, 2006; Zvolensky, Kotov, Antipova, & Schmidt, 2005). Vulnerability factors can increase the likelihood to develop an anxiety disorder and include individual differences among people, e.g., regarding personality traits or learning histories (Lonsdorf & Merz, 2017; Mineka & Zinbarg, 2006). A much-studied vulnerability factor among personality traits is subclinical trait anxiety (Raymond, Steele, & Series, 2017). High trait anxious persons describe their personality as characterized by anxiety (Taylor et al., 2007) and report a higher frequency with which anxiety symptoms are experienced (Laux, Glanzmann, Schaffner, & Spielberger, 1981). These individuals are at higher risk to develop an anxiety disorder in the future (Raymond et al., 2017). Individual differences in fear learning are another risk factor of anxiety disorders. Vulnerability can be increased by early aversive learning experiences (Mineka & Oehlberg, 2008; Mineka & Zinbarg, 2006), including differences in sensitivity or proneness to fear learning (Beckers, Krypotos, Boddez, Effting, & Kindt, 2013).

In summary, fear is an adaptive emotion (Davis et al., 2010; Lang & Bradley, 2010) that enables us to predict or respond to imminent threat. For example, it can be measured at physiological (e.g., SCR, startle) and verbal (e.g., cognitive and affective ratings) levels (Bradley & Lang, 2000; Lang, 1968). However, patients with an anxiety disorder experience fear excessively, i.e., it has become maladaptive (American Psychiatric Association, 2013; Mineka & Zinbarg, 2006), and often meet promoting

vulnerability factors, e.g., individual differences in trait anxiety or fear learning (Lonsdorf & Merz, 2017; Mineka & Zinbarg, 2006).

1.2 Conditioning model of fear

To our advantage, we do not need to evaluate every potentially threatening situation to decide on a reaction. Instead, we can refer to prior learning experiences, e.g. classically conditioned fear memories (Onat & Buchel, 2015). This helps us to spot warning signals for an imminent threat (Beckers et al., 2013). However, if retrieved unnecessarily and excessively, classically conditioned fear becomes maladaptive and promotes the development and maintenance of anxiety disorders (Lissek et al., 2005; Mineka & Zinbarg, 2006). During classical fear conditioning, a neutral stimulus is repeatedly paired with an aversive unconditioned stimulus (US). Therefore, the previously neutral stimulus becomes an aversive conditioned stimulus (CS+) predicting the US onset. The CS+ itself then induces fear in anticipation of the US (Lissek et al., 2005; Pavlov, 1927). In addition to learning CS-US contingency (Rescorla, 1968), also the emotional impact of the stimuli changes during fear conditioning. For example, the CS+ should be rated more unpleasant due to the aversive learning experience (Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002). Differential fear conditioning procedures additionally include another neutral stimulus, which is never paired with the US and consequently becomes a conditioned stimulus predicting safety (CS-; Duits et al., 2015; Lissek et al., 2005).

In the laboratory, differential fear conditioning paradigms have demonstrated that healthy individuals are able to discriminate threat signals (CS+) from safety signals (CS-). This applies to various outcome measures, including ratings of arousal and valence (Lueken et al., 2014; Schiele et al., 2016), ratings of US expectancy (Holt et al., 2014; Lissek et al., 2008), startle eye blink (Lissek et al., 2008) and SCR (Holt et al.,

2014; Schiele et al., 2016; Torrents-Rodas et al., 2013; Veit et al., 2002). Each of the listed studies found a significantly higher fear-related response to CS+ as compared to CS-.

For instance, the paradigm by Lissek et al. (2008) used rings of different diameters as conditioned cues. One ring was followed by an electric stimulus (unconditioned stimulus, US) in 75% of the trials (threat cue, CS+), while another ring was never paired with the US (safety cue, CS-). Successful fear acquisition of the healthy-participants sample was indicated by higher startle reflex amplitudes and higher ratings for US expectancy to the CS+ as compared with the CS-.

In other studies, participants learned to fear one picture of a neutral female face (CS+) paired with an US (e.g. aversively loud tone, painful pressure), but not another female face with neutral expression (CS-), which never was paired with the US (Holt et al., 2014; Lau et al., 2008; Schiele et al., 2016; Veit et al., 2002). For example, Schiele et al. (2016) used an adapted version of the screaming-lady paradigm developed by Lau et al. (2008), so called because of its compound US consisting of a face picture displaying a fearful expression and a loud female scream. The conditioning procedure consisted of two blocks, during which the CS+ was followed by US in five of six trials. Afterwards, CS+ vs. CS- was rated more arousing and more negative and elicited greater SCRs.

In summary, in laboratory studies also healthy individuals learn to respond with fear to stimuli (CS+) that reliably predict an aversive outcome (US), and to distinguish these threat signals from safety signals (CS-). These robust findings reflect the adaptive nature of fear conditioning (Beckers et al., 2013).

1.2.1 Neural mechanisms

Over the last decades, numerous studies have sought to fathom the neural mechanisms involved in the acquisition and expression of classically conditioned fear. From this, models could be inferred that are constantly developing. The amygdala has a central role in these models (Dunsmoor & Paz, 2015; LeDoux, 2000). In confrontation with a threat, the lateral nucleus of the amygdala becomes active. On the one hand, it projects to the central amygdala, where the expression of defensive behavioral reactions is initiated. On the other hand, the lateral amygdala is connected with the basal amygdala, and from there with the nucleus accumbens (NAcc), which controls defensive actions like avoidance (LeDoux, 2000; LeDoux & Pine, 2016; Ramirez, Moscarello, LeDoux, & Sears, 2015). Interestingly, even though many studies found amygdala activation during fear conditioning, the results vary with respect to the laterality of activation or temporal gradation (early vs. late activation; Sehlmeyer et al., 2009). In other words, human fMRI fear-conditioning experiments seem not to evoke consistent responses of the amygdala threat-detection circuitry. This could also explain, why a meta-analysis by Fullana et al. (2016) did not find a robust involvement of the amygdala. However, the meta-analysis confirmed several other brain regions to be activated or deactivated in response to conditioned threat or safety signals. Regarding functional activations, the authors particularly emphasized “the involvement of medial wall ‘cingulofrontal cortex’ regions, including the dorsal anterior cingular cortex (dACC) and bilateral anterior insular cortex (AIC)”. These major cortical input-output brain regions generate subjective emotional awareness from higher-order interoceptive feelings (AIC) and facilitate homeostatic autonomic and behavioral responses (dACC; Craig & Craig, 2009; Medford & Critchley, 2010).

Regarding functional deactivations, the meta-analysis by Fullana et al. (2016) identified changes in activity in the ventromedial prefrontal cortex (vmPFC), lateral orbitofrontal cortex (OFC), hippocampus and posterior cingulate cortex. Deactivations of vmPFC and OFC in the presence of a CS+ suggest that threat and safety are differentiated there (Fullana et al., 2016). Indeed, during safety signal, e.g. CS-, processing (Harrison et al., 2017) and during extinction learning and recall the vmPFC regulates the amygdala activity down so that fear expression is inhibited (Milad & Quirk, 2012). Extinction is defined as the repeated presentation of a CS+ in the absence of the US, which leads to a gradual diminishing conditioned fear response. Importantly, the fear memory is not erased, but rather a new inhibitory memory is formed, which reduces the fear of the CS.

The hippocampal complex (including dentate gyrus) is involved in distinguishing threat from safety through pattern separation processes. New memory representations are formed in the dentate gyrus in order to minimize the overlap of previous memories (Dunsmoor & Paz, 2015; Yassa & Stark, 2011).

The posterior cingulate cortex (PCC) receives its major inputs from parietal cortex and is connected with the hippocampal memory system. Thereby, the PCC is involved in spatial processing, action in space and is implicated in memory (Rolls, 2019; Vogt, 2009; Vogt & Laureys, 2009). Interestingly, PCC also receives projections from orbitofrontal cortex (OFC; Vogt & Laureys, 2009). The authors of the meta-analysis also suggest an alternative explanation for the greater PCC/hippocampal activation to the CS− representing a neural correlate of ‘relief’ related to the US omission (see also Leknes, Lee, Berna, Andersson, & Tracey, 2011).

Thank to meta-analysis, a robust pattern of activations and deactivations involved in threat processing could be identified. Nevertheless, the models will continue to evolve and become more accurate.

1.2.2 Pathologically relevant responses in fear conditioning

As already mentioned above, individual differences in fear learning are a risk factor of anxiety disorders (Mineka & Oehlberg, 2008; Mineka & Zinbarg, 2006). In line, anxiety patients show abnormalities in fear vs. safety learning (Duits et al., 2015; Lissek et al., 2005). Applying the paradigm previously described by Lissek et al. (2008) with rings of different diameters (see section 1.2) to panic patients (Lissek et al., 2010) and GAD patients (Lissek, Kaczkurkin, et al., 2014) resulted in lower differential values of CS+ and CS- during acquisition in patients compared to healthy controls. They particularly result from higher responses to CS-, indicating impaired safety learning in patients. However, this finding was observed only at the rating level, i.e., US-expectancy ratings, and not at the physiological level, i.e., startle response.

Similar results were also found in another differential fear-conditioning paradigm of Lissek and colleagues (2009), which compared panic patients with healthy controls. During acquisition phase, two neutral object images (a bowl and a mug) served as CS+ and CS-, respectively. The CS+ image was paired with an electric stimulus as US, but never the CS-. Contrary to healthy participants, panic patients reported both higher anxiety and US expectancy and showed increased startle amplitude to CS-. No group differences were observed regarding CS+. Besides, the study included a retention test after one week, until which the observed group differences did not persist.

In line, patients with social phobia compared to healthy controls showed equally strong SCRs for CS+ and CS- in in the first block of the learning phase (Veit et al.,

2002). They presented neutral faces as CSs, of which only one was coupled with an uncomfortable pressure (US).

In other words, anxiety patients show abnormalities in fear vs. safety learning insofar as their responses to safety signals (CS-) are increased, what results in delayed safety learning. The finding was robustly observed across several studies and outcome measures (e.g., US-expectancy ratings and physiological measures) and also a meta-analysis confirmed that anxiety patients differ noticeably from healthy controls by exhibiting fear responses to safety signals (Duits et al., 2015). The most common interpretations are on the one hand that anxiety patients have an impaired ability to inhibit fear to a safety signal when they are in an aversive context, and one on the other hand that patients tend to generalize learned fear responses more easily to other neutral stimuli resembling threat cues (Duits et al., 2015).

If considered a risk factor of anxiety disorders, these abnormalities should also exist in individuals with subclinical high trait anxiety. Indeed, some studies found impaired differential fear learning in high anxious compared to low anxious individuals (Andreatta & Pauli, 2017; Gazendam, Kamphuis, & Kindt, 2013; Haddad, Pritchett, Lissek, & Lau, 2012; Indovina, Robbins, Nunez-Elizalde, Dunn, & Bishop, 2011; Sjouwerman, Scharfenort, & Lonsdorf, 2020; for a review see Lonsdorf & Merz, 2017; but see Torrents-Rodas et al., 2013). Anxious personality traits are typically measured via questionnaires, in particular the trait scale of the Spielberger State-Trait Anxiety Inventory (STAI; Dymond et al., 2015; Laux et al., 1981). Based on the questionnaire score participants can be divided in high and low anxious subgroups (e.g. Haddad et al., 2012).

Haddad et al. (2012) examined moderating effects of trait anxiety on conditioned responses. In analogy to the screaming-lady paradigm (Lau et al., 2008), one female

face picture served as CS+, which was paired with a compound US consisting of a face picture displaying a fearful expression and a loud female scream. As CS-, Haddad and colleagues chose, for example, a gray oval stimulus. At the end of all three acquisition blocks, all participants have learned to differentiate between CS+ and CS- for startle, skin conductance and fear rating measures. However, high anxious individuals reported higher fear to both stimuli throughout the whole conditioning procedure as compared with low anxious participants. Moreover, high anxious showed a delayed safety learning for startle, as they showed a differential startle response in Block 2, while low anxious already in Block 1.

Similarly, Gazendam et al. (2013) found deficient safety learning in high anxious compared to low anxious control participants. Specifically, over acquisition blocks high anxious individuals showed less decrease of the startle response and reported persistently higher levels of distress in the ratings to the male face image serving as CS- than low anxious. However, the groups did not differ with regard to another male face image (CS+) associated with an electric stimulus (US).

In contrast, one study using the ring paradigm by Lissek et al. (2008) found trait anxiety to have no influence on safety learning (Torrents-Rodas et al., 2013). As already discussed by Andreatta and Pauli (2017) this might be the case because their high anxious group scored lower on the STAI as compared with the groups of Gazendam et al. (2013) and Haddad et al. (2012).

Critically, there is still no clear picture of the impact of trait anxiety on fear vs. safety learning across the various studies, as studies found inconsistent results of which outcome measures are affected by trait anxiety (Lonsdorf & Merz, 2017) and if trait anxiety is linked to stronger CS+ (Indovina et al., 2011; Sjouwerman et al., 2020) or stronger CS- response (Andreatta & Pauli, 2017; Gazendam et al., 2013; Haddad et al.,

2012). Nevertheless, it seems that also healthy individuals with subclinical anxious personality show abnormalities in fear vs. safety learning and these can be discussed as a risk factor of anxiety disorders.

1.3 Generalization of conditioned fear

Equally important as the mechanism of stimulus discrimination is the mechanism of generalization. It is the capacity to extend previous learning to new stimuli, which physically or semantically share properties with familiar, e.g. conditioned, stimuli (Dunsmoor & Paz, 2015). Thus, stimuli never associated with a US may elicit fear or defensive responses because of their similarity with the threat signal (Dunsmoor & Murphy, 2014; Dymond et al., 2015; Lissek et al., 2008). Per se, the generalization of conditioned fear responses is an adaptive mechanism, because it prevents the encounter with unknown threats. However, exaggerated generalization may be maladaptive as too many cues are perceived as threatening, prompting a constant search for safety (Dymond et al., 2015; Lohr, Olatunji, & Sawchuk, 2007). From a clinical perspective, such overgeneralization of conditioned fear responses is of particular interest (Dymond et al., 2015; Lissek, Kaczkurkin, et al., 2014).

1.3.1 Dimensions of fear generalization

About 100 years ago, the famous "Little Albert experiment" by Watson & Rayner (1920) showed that an infant acquires fear of white rats (CS) after associating them with a loud, unpleasant noise (US), and subsequently generalizes fear response to other animals, e.g., rabbits, or even fluffy objects, e.g., a fur coat (GSs). Pavlov (1927) observed generalization of conditioned learning in dogs to sensory stimuli that were physically similar to a CS (e.g., a tone of a certain frequency), which had been paired with a US (e.g., food). The conditioned response (e.g., salivation) was not only shown

following the CS, but also other stimuli like tones of different frequencies (GSs), which were never paired directly with food. In the following decades, conditioning research was highly interested in mechanisms that shape stimulus generalization gradients, focusing mainly on generalization as a function of perceptual similarity (Guttman & Kalish, 1956). To this day, perceptual similarity is the most frequently used dimension to investigate stimulus generalization in the context of fear (Dunsmoor & Paz, 2015; Dymond et al., 2015).

A typical paradigm for the investigation of similarity-based generalization of conditioned fear responses was developed by Lissek and colleagues (2008). As described previously (see section 1.2), their paradigm used rings of different diameters as conditioned cues. During generalization phase, eight further rings (generalization stimuli, GSs) were presented apart from the CSs, ranging in size between CS+ and CS-. None of the GSs was paired with the US. During the generalization phase following successful differential fear conditioning, healthy participants showed potentiated startle responses compared to CS- to four GSs most similar to CS+, i.e., GS categories GS1 and GS2. US-expectancy ratings were increased to categories GS1 to GS3. On both physiological and rating level, the response strength gradually decreased as the GSs became less similar to the CS+. These results have been replicated using the same paradigm by Torrents-Rodas et al. (2013). Their study additionally investigated fear generalization of SCR, which was extended to category GS1 only.

Another similarity-based generalization paradigm used two photographs of female faces with neutral expression as CSs, as well as eight morphs of them as GSs (Haddad, Xu, Raeder, & Lau, 2013). The US was a loud female scream presented simultaneously with a facial picture with a fearful expression (adapted from Lau et al., 2008). Its procedure is comparable to that of Lissek et al. (2008). After successful fear

acquisition, indicated by increased risk ratings and potentiated startle responses to CS+ as compared with CS-, the healthy participant sample generalized fear on both rating and physiological level. Increased startle responses relative to CS- were found to CS+ and stimulus class GS1, while increased risk for US compared to CS- was additionally reported to class GS2. Schiele et al. (2016) adapted the screaming-lady paradigm (Haddad et al., 2013; Lau et al., 2008), reducing the number of generalization stimuli to four (ranging from CS+ to CS- in 20% steps). This version of the paradigm also provoked similarity-based fear generalization, with response strength (e.g., arousal) decreasing with decreasing similarity to the CS+, and more steeply in healthy adults compared to children. A later study by the author group (Reinhard et al., 2022) demonstrated a continuous negative relationship between age (8-17) and generalization by aggregating generalization to all stimuli into a generalization index (GI) as proposed by Lenaert, van de Ven, Kaas, and Vlaeyen (2016). The GI is defined as the sum of all GSs responses divided by the CS+ response. Because it captures individual differences in generalization in a single measure, it allows for a continuous examination of generalization in relation to other variables of interest (Lenaert et al., 2016).

In summary, paradigms that examine fear generalization to perceptually similar stimuli are well established and typically result in generalization gradients with gradually decreasing response strength as the GSs become less similar to the CS+ (Dymond et al., 2015). Remarkably, it seems to be a robust phenomenon that fear generalization on rating level is broader than on physiological level (Haddad et al., 2013; Holt et al., 2014; Lissek et al., 2008; Torrents-Rodas et al., 2013). From an evolutionary point of view, such conservative bias in conscious fear responses may increase the chance of survival. A conscious perception of a potential threat may facilitate to gather more information about the stimulus by directing attentional

resources. As soon as this information suggests the presence of a threat, the autonomic fear system becomes active (Holt et al., 2014; LeDoux, 2000).

The more complex stimuli are, the more dimensions they have, the more likely higher-order processes like conceptual knowledge and inferential reasoning will be involved in fear learning and its transfer to new situations (Dunsmoor & Murphy, 2015). That is why additional branches of fear generalization research evolved, taking also into account such more complex processes (for a review see Dunsmoor & Murphy, 2015).

For example, category-based induction leads to stronger generalization of fear if the threat stimulus was a typical representative of its semantical category (Dunsmoor & Murphy, 2014). The particularity of the used paradigm was that participants were conditioned to fear one animal category (e.g., birds, CS+) associated with an electric stimulus (US), but not another animal category (e.g., mammals, CS-) typical category. Moreover, for half of the participants typical members of the respective categories served as conditioned stimuli, while for the other half atypical category members were presented as conditioned stimuli. Both groups were successfully conditioned as indicated by higher SCR to CS+ as compared with CS-. Subsequently, fear generalization was tested for three new exemplars in the CS+ (GS+) category and for three new exemplars in the CS- (GS-) category. Results revealed that fear was more readily generalized, i.e., SCRs were also elevated to novel exemplars, if fear had been conditioned to typical representatives.

Another branch focusses on how fear generalization is driven by inference rules, which have been acquired during fear learning (Boddez, Bennett, van Esch, & Beckers, 2017; Livesey & McLaren, 2009; Marstaller, Al-Jiboury, Kemp, & Dymond, 2021; Shanks & Darby, 1998; Wong & Lovibond, 2017).

In the study by Wong and Lovibond (2017), one sample went through a differential fear-conditioning and generalization paradigm. The stimulus set consisted of nine yellow squares, each containing a black dot varying in horizontal position from the left to the right. The stimulus with the black dot in the middle of the square (Stimulus 5) was selected as CS+, while Stimulus 3 served as CS-. After completion of the generalization procedure, participants were asked to indicate which relational rule they assumed between the square stimuli and the electric stimulus (e.g., similarity, linear). Interestingly, for both US-expectancy ratings and skin conductance level (SCL) the generalization patterns differed between individuals depending on their reported rule. More precisely, the subgroup of participants, who believed similarity to predict US-contingency, showed a peaked generalization gradient. It peaked at CS+ and declined with a slight asymmetry, i.e., with higher responding to stimuli right of the CS+. In contrast, for the subgroup, who indicated a linear relation between stimuli and US, a more linear generalization gradient was observed. These gradients were characterized by peak responses to the stimulus at the right end of the continuum (Stimulus 9) and a clear step between CS+ and CS-. In conclusion, higher order processes like inferential reasoning can be included in fear generalization. Inferred rules then can lead to gradients that deviate from similarity-based gradients.

1.3.2 Neural mechanisms

Research on neural mechanisms underlying generalization of conditioned fear, so far has focused on generalization to perceptually similar stimuli (Dunsmoor & Paz, 2015; Lissek, Bradford, et al., 2014; Onat & Buchel, 2015). For instance, Lissek, Bradford, et al. (2014) used a version of the previously described paradigm with rings of different diameter. fMRI-activity data revealed that generalization was accompanied by two neural activation patterns: Firstly, a positive generalization gradient, reflected in

decreasing responses the more the stimuli differ from CS+ in bilateral anterior insula, dorsomedial prefrontal cortex (dmPFC) and bilateral inferior parietal lobule. Secondly, a negative gradient, reflected in increasing responses with increasing difference to CS+ in bilateral ventral hippocampus, ventromedial prefrontal cortex (vmPFC) and precuneus cortex. These results are consistent with a neurobiological model of fear generalization, already previously outlined by Lissek and Grillon (2012). The model proposes two pathways of fear learning and expression. One pathway directly directs information about potential threat stimuli from sensory thalamus to the amygdala, herby bypassing sensory cortex. Therefore, in the face of potentially threat-relevant generalization stimuli with strong similarity to CS+, the expression of conditioned fear can be rapidly initiated via output connections with the insula and brainstem. The other pathway involves the processing of cortical representations in the hippocampus, which plays a role in pattern completion or rather separation. The hippocampus performs a same-different determination for every cortical representation of a new stimulus event (i.e., GS) comparing it to similar representations of past events stored in memory (i.e., CS+; Kaczkurkin et al., 2017; Otto & Eichenbaum, 1992; Yassa & Stark, 2011). In case of sufficient overlap of new and stored representation, the hippocampus initiates a pattern completion process, hereby reactivating the neural representation of the CS+, which then triggers the conditioned fear response. In turn, insufficient overlap initiates the process of pattern separation, which goes hand in hand with activation of the vmPFC and consequently a down regulation of the amygdala (Lissek, Bradford, et al., 2014; Lissek & Grillon, 2012). In conclusion, this model of fear generalization suggests that fear responding on neuronal level is dependent on perceptual similarity (Lissek, Bradford, et al., 2014; Onat & Buchel, 2015).

In line with the pattern separation hypothesis, a better performance in a task assessing pattern-separation abilities, was associated with less fear generalization to

perceptually similar stimuli as reflected in US-expectancy ratings (Lange et al., 2017). For the "pattern-separation" task, participants first had to classify 128 object images presented after one another as indoor or outdoor objects. In the retrieval phase, each of the 192 images should then be identified as exactly the same, similar, or completely new. Performance of this task was related with a reduced activity of the subcallosal cortex during assessment of fear generalization, a region with connectivity to orbitofrontal cortex and vmPFC.

Generalization-relevant perceptual changes in neural processing can already occur in primary sensory regions (Laufer, Israeli, & Paz, 2016). The fMRI study compared patients with GAD to healthy controls. First, a trial-and-error task was conducted, aiming participants to learn the outcomes of three tone-key combinations, one predicting gain of money (positive tone, PT), one a zero outcome and one loss of money (negative tone, NT). In a subsequent test phase, all tone-key combinations were instructed to have no outcome. Still, participants were asked to press the previously learned key if they recognize the PT or NT, and a third key to all further tones. As expected, GAD patients showed a wider behavioral generalization to new tones with similarity to PT and NT. Imaging revealed a correlation of amygdala activity and behavioral generalization, again for both gain and loss. Importantly, the activity of auditory cortex indicates less neural discrimination between CS (NT or PT) and GSs in patients compared to healthy controls. These findings suggest that the affective modulation of stimulus presentations in primary cortices and amygdala contributes to fear generalization. Consequently, overgeneralization in anxiety patients might have perceptual origins.

In contrast, a study by Onat and Buchel (2015) revealed that ambiguity-based uncertainty evaluations are involved in fear generalization, suggesting that active

processes contribute to fear generalization. They designed a new fear-generalization paradigm, whose face stimuli formed a circular continuum of perceptual similarity, making it possible to examine fear generalization as similarity-determined vs. ambiguity-related active process. In the first place, results confirmed that fMRI activity in hippocampus and higher cortical areas (i.e., vmPFC) followed perceptual similarity, as did the behavioral generalization gradient. Besides, the involvement of the anterior insula was confirmed. However, its activation pattern was hypersharp indicating less generalization as compared with the behavioral gradient. This discrepancy can be explained through the activation of the ventral inferior temporal cortex (ITC), typically involved in perceptual processing, which exhibited fear responses based on ambiguity-related uncertainty. The integration of insular and ITC responses determined the sharpening of behavior. This view suggests that fear generalization arises as a consequence of active processes.

In summary, research on neural mechanisms of fear generalization found evidence for both perceptual (Laufer et al., 2016; Lissek, Bradford, et al., 2014) and higher-order processes (Onat, 2018; Onat & Buchel, 2015) to be involved in fear generalization. Further neuroimaging studies on generalization are warranted to better understand underlying mechanisms and their interplay.

1.3.3 Pathologically relevant responses in similarity-based fear generalization

In a series of studies, Lissek and colleagues demonstrated that anxiety patients (Lissek, Kaczkurkin, et al., 2014; Lissek et al., 2010) exhibit stronger similarity-based generalization of conditioned fear responses as compared with healthy controls. The application of the previously described paradigm with rings of different diameter by Lissek et al. (2008) (see sections 1.2 and 1.3.1) to panic patients (Lissek et al., 2010)

and GAD patients (Lissek, Kaczkurkin, et al., 2014) revealed that panic patients showed fear-potentiated startle responses to a broader range of GSs and believed the US to be more likely after all stimuli as compared to healthy individuals. Similarly, GAD patients had more problems to distinguish GS1 from CS+, as the CS+/GS1 difference of both startle amplitude and US expectancy was smaller in patients as compared with healthy controls (but see Tinoco-Gonzalez et al., 2015).

Overgeneralization of conditioned responses to perceptually similar auditory stimuli was supported by the work of Laufer et al. (2016). They compared patients with generalized anxiety disorder with healthy people. As described in section 1.3.2, during the test phase, participants were asked to press the learned key when they recognized a previously conditioned positive or negative tone predicting a monetary gain or loss, respectively. For all further generalization tones they were asked to press a third key. Importantly, during test phase, none of the tone-key combinations caused monetary gain or loss. It is noticeable that anxiety patients over-generalized the loss-avoidance reactions to significantly more new tones than healthy people, even though they could no longer benefit from them because the tone-key combinations no longer had any consequence. Interestingly, similar results were found for the generalization of gain reactions, contradicting a “better safe than sorry” decision rule. The authors argue that the affective stimuli formed early sensory representations during conditioning, which later led to a poorer discrimination between conditioning stimuli and new safe stimuli.

In summary, anxiety patients show stronger fear responses and to a broader range of GSs, which were not associated with the US but resemble the CS+, as compared to healthy individuals. Contrary to healthy controls, anxiety patients seem to over-generalize their conditioned fear (for a meta-analysis see Cooper et al., 2022), and

this is discussed as a pathogenic marker of anxiety disorders (Lissek & Grillon, 2010; Lissek et al., 2010).

The relation between vulnerability to the development of clinical disorders and fear generalization is also influenced by other risk factors, e.g. trait anxiety (Raymond et al., 2017). In their review on computational models to study risk for anxiety disorders, Raymond et al. (2017) suggest that individuals with a more anxious predisposition (e.g. trait-anxiety) show more generalized aversive learning and behavior, and consequently experience more general anxiety. The resulting increase in frequency of anxious states then intensifies the vulnerability for an anxiety disorder.

In line, several studies revealed that healthy individuals with high trait anxiety show a tendency to over-generalize conditioned fear too (Baumann et al., 2017; Haddad et al., 2012; but see Torrents-Rodas et al., 2013).

For instance, Haddad et al. (2012) examined how trait anxiety generalizes to a safety cue (CS-), which was perceptually similar to the CS+. To this end, they developed a novel fear-conditioning paradigm using three conditioned stimuli. As mentioned in section 1.2.2, one female face picture served as CS+ and a grey oval stimulus as (dissimilar) CS-. Another face picture with perceptual similarity with CS+ was presented as similar CS-. All participants generalized the conditioned fear to the perceptually similar CS- for startle, skin conductance and fear rating measures. Regarding startle, this fear generalization was particularly strong in the high anxious compared to the low anxious group. More precisely, during Acquisition Block 2, the startle response of high anxious individuals to the similar CS- did not differ from CS+ reaction. In contrast, low anxious persons seem to discriminate the stimuli better, as their startle to similar CS- was lower as compared with CS+. However, these group differences did not persist during Acquisition Block 3.

A relation between anxious personality and fear generalization was also supported by studies, which determined generalization of fear from an aversive conditioned face picture (CS+) to four morph stimuli (GS), ranging in perceptually similarity from CS+ to CS- in 20 % steps (Baumann et al., 2017; Stegmann et al., 2019; for paradigm also see Schiele et al., 2016). With a principal component analysis of several questionnaires Baumann et al. (2017) first identified a factor representing the construct anxiety, characterized by the Anxiety Sensitivity Index (ASI 3; Taylor et al., 2007), the Agoraphobic Cognitions Questionnaire (ACQ; Chambless, Caputo, Bright, & Gallagher, 1984) and the Social Phobia and Anxiety Inventory (SPAI; Turner, Beidel, Dancu, & Stanley, 1989). In order to compare low and high anxious participants with regard to fear generalization, two extreme groups of anxiety-factor values were separated, with the low 25 % quartile values forming the low anxious group, and the high 75% quartile values forming the high anxious group. As a result of membership in one extreme group, fear generalization of valence varied. While both low and high anxious groups reported more negative valence to CS+ and the GS most similar to CS+ (GS1) as compared with CS-, only the high anxious group additionally reported more negative valence to GS2. Similarly, exploratory analysis of US-expectancy ratings revealed that both groups perceived a higher risk for the US compared to CS- for CS+, GS1 and GS2, high anxious individuals, however, for GS3 and GS4 as well.

The study by Stegmann et al. (2019) found some evidence for anxious personality being associated with specific patterns of fear generalization. Individuals with a generalization pattern, which is characterized by lower average fear response and good stimulus differentiation (causing a steep gradient pattern), scored lower on questionnaires related with anxiety (e.g., STAI trait, ASI 3, ACQ, SPAI, Liebowitz Social Anxiety Scale LSAS; Stangier & Heidenreich, 2003) compared to individuals

with generalization patterns characterized by high average fear response and flat gradients.

In summary, despite predominantly small effects of trait anxiety on fear generalization and existing contradicting findings (e.g., Torrents-Rodas et al., 2013), a meta-analysis confirmed that high levels of anxious personality characteristics in healthy individuals are associated with stronger generalization of fear (Sep, Steenmeijer, & Kennis, 2019). Importantly, this meta-analytic finding supports the assumption that fear generalization is an existing mechanism already before onset of an anxiety disorder and therefore can be considered a risk factor for pathological anxiety.

1.3.4 Perception-dependent generalization of fear

As mentioned above, conditioned fear is generalized to perceptually similar cues and this mechanism is exaggerated in anxiety patients (Dunsmoor & Paz, 2015; Dymond et al., 2015). Following up on these findings, some studies addressed the question if fear generalization occurs as consequence of an incapacity in perceptually discriminating stimuli (Holt et al., 2014; Struyf, Zaman, Hermans, & Vervliet, 2017; Zaman, Ceulemans, Hermans, & Beckers, 2019).

In that regard, it also is of interest if and how aversive learning alters perceptual thresholds and hereby promotes generalization (Resnik, Sobel, & Paz, 2011; Schechtman, Laufer, & Paz, 2010; Shalev, Paz, & Avidan, 2018). Accordingly, Shalev et al. (2018) in a series of studies found that aversive conditioning deteriorates stimulus discrimination. Specifically, the authors compared participants' stimulus discrimination before and after aversive (experimental group) or neutral conditioning (control group). During discrimination task, either CS+ or CS- and a second stimulus were presented in a random order. The two stimuli differed only slightly on one perceptual dimension (e.g., tone frequency). Participants had to indicate which of the two stimuli had the

larger magnitude on the dimension (e.g., higher tone frequency). Remarkably, after aversive but not neutral conditioning the discrimination threshold was increased compared to baseline threshold.

Holt and colleagues (2014) examined fear generalization to stimuli below or above perceptual thresholds. To create GSs below an individual's perceptual threshold, individual just noticeable differences (JND) for face stimuli were assessed with a discrimination task, consisting of three blocks containing 50 trials, each. Every task trial began with the presentation of the future CS+ face for 500 ms. Following a 500 ms inter-stimulus-interval (ISI), two stimuli appeared on the screen, both the CS+ and one of the possible GSs morphs or CS- face. Participants' task was to indicate by pressing a key whether the left or right picture was identical to the picture presented first. Participants could respond without time limit. Trials were separated by an inter-trial-interval (ITI) of 1 s. JND was the morph level (% difference from CS+) which could be correctly discriminated from CS+ with 75% accuracy. Three GSs were set below individual JND, two GSs at and above JND, respectively. Interestingly, after successful discriminative fear learning, in the generalization test, participants reported increased risk for the US, i.e., generalized fear, to all GSs below and at JND, but not above JND. SCR was only increased to two GSs, both below JND. Consequently, fear seems not to generalize to stimuli which can be reliably discriminated.

In line, some studies combined a fear conditioning and generalization experiment with a perceptual decision task and found evidence that generalization to perceptually similar stimuli is driven by perceptual errors, i.e., misidentification of a novel stimulus as the conditioned fear stimulus (Struyf et al., 2017; Zaman et al., 2019). For example, the stimulus set of the conditioning and generalization experiment by Zaman et al. (2019) consisted of seven white circles of different sizes, of which the

middle circle was assigned to the function of the CS paired with an aversive IAPS picture as US. All further circles served as GSs and were never paired with the US. During the generalization phase, participants were asked to identify each presented stimulus (i.e., CS or GS) as identical or different from the stimulus presented in the previous conditioning phase (i.e., CS). After the perceptual decision, each stimulus was evaluated for US expectancy. The results showed that GSs, which had been incorrectly identified as CS, were rated with a higher US expectancy. Moreover, individual variation in perceptual errors could be clustered in perceptual discrimination patterns. Discrimination patterns characterized by many perceptual errors across GSs or misidentification of extreme GSs on one side of the continuum were associated with higher US expectancies to GSs. In contrast, the discrimination pattern characterized by sharp discrimination of CS led to decreasing US expectancy as a function of perceptual similarity. In summary, the results suggest that generalization to perceptually similar stimuli depends on perception as it determines behavior directly but also indirectly via modification of learning experiences.

In summary, several studies found evidence for perception-dependent fear generalization, as also summarized in a review by Zaman et al. (2021), in which the authors recommend the inclusion of perceptual measures in generalization research, as this can explain both intra- and interindividual differences in generalization.

1.3.5 Reducing similarity-based fear generalization by discrimination trainings

According to the findings summarized in the previous section, improving cue discrimination should be a promising approach to reduce fear generalization. Following this idea, some studies developed discrimination trainings in order to prevent fear generalization (Ginat-Frolich, Gendler, Marzan, Tsuk, & Shechner, 2019; Ginat-

Frolich, Klein, Katz, & Shechner, 2017; Lommen et al., 2017). Specifically, Ginat-Frohlich and colleagues (2017, 2019) realized a discrimination training with visual stimuli which were not related to those presented during fear conditioning or assessment of generalization. The discrimination training was applied after fear acquisition and consisted of 63 trials. Each trial was structured as follows. First, one abstract shape (target) was presented for 4 s in the middle of a computer screen. After a 2 s fixation point, the target as well as a new similar shape were presented on the left and on the right of the computer screen. Participants should identify the target as fast and accurately as possible. Participants who had been assigned to the non-discriminative control task were shown the same shapes, but only one at a time, either on the left or on the right side of the screen. Their task was to indicate on which side the shape appeared. In both tasks, participants received visual feedback on their choice. After completion of the discrimination training or control task, participants underwent a generalization test. In the beginning of the experiment, participants had successfully learned to fear an image of a bell with a certain color (e.g., blue, CS+), which had been paired with a 95 dB loud aversive sound (US) in 80% of presentations. A bell of another color (e.g., yellow) had never been paired with the US and served as CS-. The generalization test assessed to which extent participant generalized the conditioned fear response to nine differently colored bell morphs (GSs) ranging in perceptual similarity from CS+ to CS-. In both studies (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017), the discrimination-training group showed less fear generalization to the differently colored GS-bells than the control-task group. However, in healthy adults this effect was restricted to US-expectancy ratings and was not found for startle response (Ginat-Frolich et al., 2017), while in the children sample it was vice versa (Ginat-Frolich, Gendler, et al., 2019). Nevertheless, the discrimination training (as compared to a non-

discriminative control task) improved stimulus discrimination and consequently reduced generalization.

Another working group developed a discrimination training whose stimuli shared one perceptual characteristic (i.e., color) with the conditioned stimuli (Lommen et al., 2017). In each of the 30 trials, two stimuli were presented shortly after each other. Participants had to decide if the stimuli differed in size or color or not at all. Half of the participants completed this discrimination training, while the other half was assigned to a non-discriminative control task. In trials of the control task, only one stimulus was presented and participants had to indicate if the word in the left or right bottom corner of the screen related to this stimulus. The words related to either the shape (e.g., trapezium, circle) or the color (e.g., white, elephant) of the stimulus. Both tasks included audio feedback about correctness of the participant's decisions. After completion of the discrimination training or control task, participants underwent a differential fear conditioning and generalization procedure using a stimulus set of ten triangles on a spectrum from white to black. Unlike with standard differential conditioning paradigms, two stimuli, i.e., the two lightest and two darkest triangles, served as CSs+ and CSs-, respectively. CSs+ were reinforced with an electric stimulus US in 100% of presentations. During generalization phase, each of the ten triangles was presented twice. Only CSs+ were followed by US in 100% of presentations. Fear acquisition was successful, as both groups reported higher US-expectancy ratings to CSs+ than CSs-. Contrary to the training of Ginat-Frolich et al. (2017), this discrimination training did not reduce fear generalization of US expectancy compared to the control-task group. However, the discrimination-training group showed less avoidance behavior to the GSs as compared with the non-discriminative control task group.

In summary, these studies suggest that discrimination trainings with either fear-irrelevant or partial fear-relevant stimuli reduce the occurrence of fear generalization.

1.4 Research question

As summarized in the previous sections, the generalization of learned fear to novel cues that resemble a familiar threat is an adaptive mechanism, but one that is pathologically exaggerated in anxiety patients (Dymond et al., 2015). Generalization of conditioned fear has been studied particularly on the dimension of perceptual similarity (Dunsmoor & Paz, 2015; Dymond et al., 2015). It appears to be related to an incapacity to perceptually discriminate threat from similar safe stimuli (Holt et al., 2014; Zaman et al., 2019). Consequently, improving stimulus discrimination is a promising approach to reduce fear generalization. Indeed, some studies have developed discrimination training that successfully reduced the occurrence of fear generalization (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017; Lommen et al., 2017). However, it is important to point out that these trainings were conducted before (Lommen et al., 2017) or after (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017) fear learning. Therefore, only the training in the latter studies can be considered treatment, while the former study represents prevention training. Furthermore, in the aforementioned studies, discrimination training was conducted before participants' generalization gradients were examined. This is a limitation because the authors could only examine whether discrimination training prevents or at least reduces the occurrence of fear generalization compared to a control group. However, patients with anxiety disorders seek treatment after acquiring fear and with a long history of overgeneralization. This raises the question of whether the use of discrimination training both after fear acquisition and after an initial demonstration of fear generalization can reduce the extent of

generalization. This approach mimics a therapeutic intervention and will be examined for the first time in this dissertation.

The importance of this question stems from the fact that overgeneralization of learned fear is considered a risk factor for the development of anxiety disorders (Lissek & Grillon, 2010). Other risk factors include individual differences in, for example, earlier learning experiences (e.g., abnormalities in fear vs. safety learning) or anxious personality trait (Lonsdorf & Merz, 2017; Mineka & Oehlberg, 2008; Mineka & Zinbarg, 2006). Interestingly, both abnormalities in fear vs. safety learning (Gazendam et al., 2013; Haddad et al., 2012) and a tendency to over-generalize (Sep et al., 2019) are already present in highly anxious individuals, supporting the notion that these are personality traits that make the disorder more likely (Raymond et al., 2017; Sep et al., 2019). Therefore, the second study in this dissertation will examine how these individual differences influence fear generalization as well as its potential reduction through discrimination training.

2 Study 1: Reducing evolved generalization of conditioned fear through discrimination training – a proof-of-principle study

The aim of the first study of this doctoral thesis was the development of a discrimination training that reduces fear generalization after it has already occurred. One part of this study (Study 1A) has been published at Frontiers of Psychology (K. Herzog et al., 2021). Parts of Study 1A were topic of the master thesis by Paula Engelke, which was prepared under my supervision.

2.1 Introduction

The research idea of Study 1 arose from findings that anxiety patients (Laufer et al., 2016; Lissek et al., 2010) and people at risk for an anxiety disorder (Baumann et al., 2017; Haddad et al., 2012) tend to over-generalize fear to perceptually similar stimuli. Some studies already put this idea into practice by carrying out preventive discrimination training (Lommen et al., 2017) or training after fear-learning (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017). Following up on their results, this study aims to address unanswered and further-reaching questions.

The first and possibly most important question is whether a discrimination training cannot only prevent fear generalization but also reduce it after occurrence. As already mentioned in the previous chapter, anxiety disorder patients are seeking for treatments after fear acquisition and with a long history of overgeneralization. For this reason, it is important to expand research on trainings that mimic a therapeutic intervention. Accordingly, in this study, participants completed the training after fear acquisition and a first demonstration of fear generalization.

Another unanswered question is whether fear-relevant trainings are more effective than fear-irrelevant ones. Interestingly, the fear-irrelevant discrimination

training used by Ginat-Frolich and colleagues (2019; 2017) reduced generalization for US expectancy. Based on these studies it seems that the simple performance of a discrimination training (with any stimuli) should be enough to reduce generalization processes. However, this was not supported by the findings of Lommen and colleagues (2017), who applied a training with partial fear-relevant stimuli, which failed to reduce generalization of US expectancy. Besides, there is evidence that perceptual learning cannot be easily transferred to an unpracticed stimulus set (Furmanski & Engel, 2000), suggesting a higher efficacy of trainings with fear-relevant stimuli. Therefore, in this study the issue was examined further by comparing the efficacy of discrimination trainings with fear-relevant vs. fear-irrelevant stimuli (Study 1A). In a second part of the study (Study 1B), the discrimination trainings were compared with a control task, which did not entail perceptual discrimination.

A third aspect I was interested in in Study 1 is the role of feedback during discrimination training. Both the studies by Ginat-Frolich and colleagues (2019; 2017) and Lommen and colleagues (2017) included feedback during discrimination training. Other studies demonstrated that positive feedback is able to improve stimulus perception (Sasaki, Nanez, & Watanabe, 2010), particularly when associated with reward (Weil et al., 2010). That is why I examined whether rewarding feedback on discrimination performance further enhances training efficiency compared to training without feedback. In Study 1B, I was able to particularly address the question if simply giving rewarding feedback is sufficient to reduce fear generalization.

2.2 Study 1A: Fear-reducing effect of fear relevance and feedback in discrimination training

Study 1A was designed to examine whether discrimination training can reduce fear generalization after occurrence and what role fear relevance or feedback play during training. For this purpose, all participants underwent a well-established differential fear conditioning and generalization paradigm (Baumann et al., 2017; Schiele et al., 2016). The discrimination training occurred between the two generalization blocks. One half of participants trained to discriminate fear-relevant stimuli, i.e., discrimination of the conditioned threat stimulus (CS+) from generalization stimuli (GSs) and the conditioned safety stimulus (CS-). The other half was assigned to a training with fear-irrelevant stimuli, i.e., discrimination of a specific line width from other lines. Both trainings occurred either with or without feedback on discrimination performance. I hypothesized that the fear-relevant training reduces fear generalization more effectively than the fear-irrelevant training, and that this effect is especially strong if feedback is given.

2.2.1 Materials and methods

2.2.1.1 Participants

The study was advertised on an Internet platform of the University of Würzburg (psywue.sona-systems.com). Every person who was interested in participation was screened for exclusion criteria by a telephone interview. Exclusion criteria were psychiatric and neurological disorders, intake of psychoactive medication, excessive consumption of alcohol or nicotine and pregnancy. Moreover, only participants between the ages of 18 and 50 were included. In the end, 80 participants were invited to the laboratory. After arrival, all participants read and signed the informed consent and then were randomly assigned to one of four experimental groups. Group descriptions are

summarized in **Table 1**. I found no pre-experimental differences between the groups. For a broader sample description, participants completed the German version of the State-Trait Anxiety Inventory (STAI; Laux et al., 1981) and the Beck-Depression Inventory (BDI II; Hautzinger, Keller, & Kühner, 2006) before the experiment (**Table 1**). The study was approved by the ethical committee of the medical board of the University of Würzburg and was conducted in accordance with the ethical principles of the Helsinki Declaration.

Table 1 *Sample characteristics of Study 1A*

	relevant_DT_noFB	*relevant_DT_FB*	*irrelevant_DT_noFB*	*irrelevant_DT_FB*	*comparisons*
N	20	20	20	20	
gender (♀)	10	10	10	10	
age (*SD*)	24.25 (4.27)	26.10 (8.42)	25.70 (7.40)	24.75 (5.37)	$F(1,76) = 0.91$, $p = .344$
language (german)	17	17	18	19	$\chi(3) = 2.16$, $p = .540$
handedness (right)	19	19	19	17	$\chi(1) = 0.72$, $p = .396$
STAI (*SD*)	34.30 (9.14)	32.70 (6.53)	34.90 (10.18)	35.25 (7.41)	$F(1,76) = 0.27$, $p = .607$
BDI (*SD*)	5.1 (5.91)	3.70 (4.46)	5.25 (6.21)	5.55 (6.35)	$F(1,76) = 0.43$, $p = .513$

Notes. DT discrimination training, *FB* feedback, *STAI* State-Trait-Anxiety-Inventory Trait Scale, *BDI* Beck Depression Inventory II.

2.2.1.2 Questionnaires

During the experiment, the participants had to complete a series of questionnaires. On the one hand, demographic information was requested, e.g., age,

gender, native language, and handedness, and on the other hand, the German versions of some questionnaires.

Depressive symptomatology (e.g., depressed mood, sleep disturbances, suicidality) was assessed with the Beck Depression Inventory (BDI II; Hautzinger et al., 2006). The 21 items can take a value from zero to three in steps of one. The sum value indicates the severity of depression.

To measure a general tendency toward anxious reactions as a personality trait, participants completed the trait questionnaire of the State-Trait Anxiety Inventory (STAI; Laux et al., 1981). It includes 20 items on a 4-point Likert scale ranging from one (“almost never”) to four (“almost always”), which are summed to a total score.

In addition, the State version of the STAI (Laux et al., 1981) and the Positive and Negative Affect Schedule (PANAS; Krohne, Egloff, Kohmann, & Tausch, 1996) were completed for assessment of emotional state at the beginning and at the end of the experimental protocol (for details see Appendix 6.1.1). Like the trait version, the State Scale of the STAI contains 20 items that are answered on a 4-point Likert scale and summed to produce a total score. The PANAS captures participants' mood and consists of 20 items, 10 of which characterize positive (e.g., active) and 10 negative (e.g., distressed) sensations and feelings. Their presence is indicated on a 5-point Likert scale ranging from one ("very slightly or not at all") to five ("extremely"), each of which adds up to a sum score for positive and negative mood.

2.2.1.3 Stimulus material

Two pictures showing neutral face expressions of a brunette and a blond woman (03F_NE_C, 10F_NE_C, NimStim Face Stimulus Set, Tottenham et al., 2009) were used as conditioned stimuli (CS). One face stimulus (CS+) was paired with the aversive unconditioned stimulus (US) in 83% of the trials. In contrast, the other face (CS-) was

never paired with the US (for procedure details also see section 2.2.1.4). The assignment of the faces to CS+ and CS- function was counter-balanced across participants.

The US was a compound stimulus, consisting of a 95 dB female desperate scream (International Affective Digitized Sounds, IADS, FemScream2, No. 276; Bradley & Lang, 1999) and a picture of the CS+ face with fearful expression, presented simultaneously at the offset of CS+ for 1.5 s.

Four generalization stimuli (GSs) were created by gradually morphing the CSs in 20% steps using the graphics software Squirlz Morph (for details see Schiele et al., 2016, Version 2.1, Xiberpix, Solihull, UK). CSs and GSs were presented for 6 s each.

At the end of every experimental block (also see section 2.2.1.4), participants were asked to rate each face stimulus on arousal ("how much stress/tension/arousal was elicited by this stimulus?") and valence ("how pleasant vs. unpleasant was the stimulus for you?") on Likert scales from 1 ("calm" or "very unpleasant") to 9 ("intense" or "pleasant"). Then, US-expectancy ratings ("how high is the probability that you will hear the scream by this stimulus?") were obtained using a Likert scale ranging from 0 ("very unlikely") to 100 ("very likely"), except after habituation phase. During each rating block, the faces were rated in the same order (brunette woman, blond woman, morphs from brunette to blond) and were presented for 1 s each before a Likert scale appeared.

2.2.1.4 Procedure

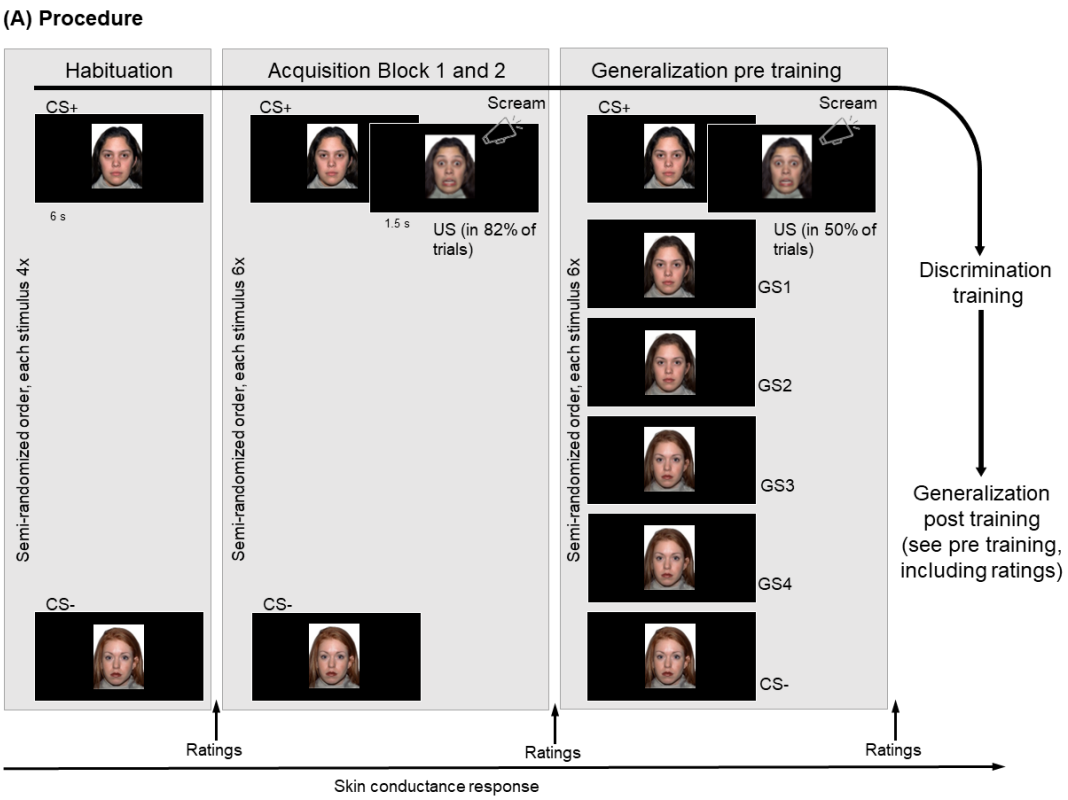
After arrival at the laboratory, participants were seated in a comfortable chair, gave their informed consent for study participation and filled in the questionnaires. Then, the electrodes for physiological recordings and the headphones for US presentation were attached. The experiment was conducted using Presentation software

Version 16.0 (Neurobehavioral Systems, Inc., Albany, CA). Participants were instructed to passively view pictures and that they will occasionally hear an unpleasant loud sound. The instructions did not include information about the CS-US-contingency.

The experimental procedure (**Figure 1** A) was based on previous studies (Lau et al., 2008; Schiele et al., 2016). Stimuli were presented in a pseudo-randomized order so that the same stimulus appeared not more than twice in a row. The time between two stimulus presentations (inter-trial-interval, ITI) ranged between 9 and 12 s randomly, during which a white fixation cross was displayed in the center of the screen.

During habituation phase, both the CS+ and the CS- were presented four times each without any US. The acquisition phase and generalization phase were divided into two identical blocks, containing six presentations of each stimulus. In acquisition blocks, the CS+ was followed by US in five trials out of six (i.e., 83% reinforcement), while the CS- was never followed by US. During generalization blocks, i.e., generalization tests, all faces were presented and three CS+ trials were followed by US to prevent extinction of conditioned fear. Between the generalization blocks, participants underwent discrimination trainings (see section 0).

At the end of each block, participants were asked to rate arousal and valence of the faces. Moreover, they indicated their US expectancy (except after habituation phase). Throughout the whole experiment, skin conductance responses (SCRs) were recorded. After the experiment participants again completed the state questionnaires.


(A) Procedure
Habituation
Acquisition Block 1 and 2
Generalization pre training
CS+
Scream
6 s
1.5 s
US (in 82% of trials)
US (in 50% of trials)
Semi-randomized order, each stimulus 4x
Semi-randomized order, each stimulus 6x
Semi-randomized order, each stimulus 6x
GS1
GS2
GS3
GS4
CS-
Discrimination training
Generalization post training (see pre training, including ratings)
Ratings
Ratings
Ratings
Skin conductance response


(B) Fear-relevant discrimination training
500 ms
500 ms
Is this picture identical to the previous one?
participants' response
no-feedback group
1 s
feedback group


(C) Fear-irrelevant discrimination training
500 ms
500 ms
Is this line identical to the previous one?
participants' response
no-feedback group
1 s
feedback group

Figure 1. Overview of the experimental procedure.

Notes. During the experiment (A), participants repeatedly saw two faces with neutral expressions. In the acquisition blocks, one face (CS+), but not the other face (CS-), was followed by a compound unconditioned stimulus (US) consisting of a loud scream and the CS+ face with a fearful expression. In the generalization phase, participants additionally saw morphs between the two initial faces (GSs). Discrimination training was embedded in two generalization blocks, and participants were divided into four groups: Two groups underwent fear-relevant discrimination training (B), and the other two groups underwent fear-irrelevant discrimination training (C), each either with or without feedback on discrimination performance. For copyright reasons, the face images shown here differ from the images used.

2.2.1.5 Discrimination trainings

The training protocol was inspired by the discrimination task described previously (Holt et al., 2014). The number of presentations of each face was hold constant across all trainings.

For the fear-relevant discrimination training (**Figure 1** B), every trial consisted of three elements: a 500 ms CS+ presentation, followed by a 500 ms inter-stimulus interval (ISI), and another stimulus presentation. The second stimulus was either CS+, one GS, or CS-, and was displayed above the question “Is this picture identical with the previous one?”. Participants were asked to respond by pressing the S and L on a German keyboard, which were assigned to the answers yes or no respectively. Response time was not limited. Therefore, face and question lasted on the screen until response was given.

In one group of participants (relevant_DT_FB), correct answers were reinforced by feedback, consisting of a 10 Euro cent picture and a cash register sound (52.4 dB) lasting 1 s. Incorrect answers were not followed by feedback. In another group (relevant_DT_noFB), no feedback was provided at any time.

The protocol of the fear-irrelevant discrimination training (**Figure 1** C) was the same as described for the fear-relevant discrimination training. However, all faces were

turned upside down and a blue line (35-79 pixel wide) was presented on the right or left side. Importantly, the first stimulus-line had always the smallest width and was always combined with the upside-down CS+ face. Participants had to answer the question whether the width of the lines was identical or not. Again, one group (irrelevant_DT_FB) received feedback when answering correctly, but not the other group (irrelevant_DT_noFB).

Altogether, participants performed 50 training trials with a short break after 25 trials. In all trainings, trials were separated by an ITI lasting 1-3 s randomly. For 20 out of the 50 trials the two pictures were identical, as the second stimulus was the CS+ or the thinnest line. For the remaining 30 trials, the two pictures were not identical, each of the remaining five faces or lines were presented six times. While the thin line of the first stimulus was always combined with the upside-down CS+, the line-face combinations of the second stimulus varied. Nevertheless, as during the fear-relevant training, in the fear-irrelevant training every non-CS+ face was presented six times each.

Due to a programming error, I cannot report training performance of the groups without feedback, but training performance of the groups with feedback only.

2.2.1.6 Data recording and reduction

Skin conductance responses (SCRs) were recorded continuously throughout the experiment with two 8 mm Ag/AgCl electrodes attached at the thenar and hypothenar eminences of the non-dominant hand. For recording, Brainproducts V-Amp and BrainVision Recorder software (Version 1.21, Brainproducts, Gilching, Germany) were used. Sampling rate was set to 1000 Hz and an online notch filter removed frequencies of 50 Hz. Offline analyses were run with BrainVision Analyzer Software (Version 2.1, Brainproducts, Gilching, Germany). First, a high cutoff filter removed frequencies above 1 Hz. In accordance with the guidelines (Boucsein et al., 2012), SCRs were

defined as the difference in µS between response onset (900-4000 ms after stimulus onset) and peak (2000-6000 ms after stimulus onset). A minimum response criterion of 0.02 µS was applied, with lower responses scored as 0. Next, every reaction of each participant was range corrected by dividing it by the individual's strongest reaction to a face picture (CS or GS). Besides, all SCRs were transformed into $\log_{10}(SCR + 1)$. Then, values were averaged for each stimulus and experimental block. Participants ($n = 5$) with an overall raw mean response smaller 0.02 µS were defined as non-responders and therefore excluded from statistical testing of SCR. Accordingly, $n = 20$ of the relevant_DT_FB group, $n = 16$ of the relevant_DT_noFB group, $n = 19$ of the irrelevant_DT_FB group, and $n = 20$ of the irrelevant_DT_noFB group were included in statistical analysis of SCR.

2.2.1.7 Statistical analysis

Statistical analyses were carried out in the R software environment (version 3.6.1) using packages 'afex' (version 0.26-0; Singmann, Bolker, Westfall, & Aust, 2015) and 'emmeans' (version 1.4.5; Lenth & Lenth, 2018).

Effects of fear acquisition were analyzed with ANOVAs. They presented the within-subject factors stimulus (CS+, CS-) and block, with the factor block having three levels (habituation, acquisition 1, acquisition 2) for arousal, valence and SCR data, and two levels (acquisition 1, acquisition 2) for US-expectancy data.

In order to investigate discrimination-training effects, I first calculated a Generalization Index (GI) for each generalization block defined as the sum of all GSs responses divided by the CS+ response (GI = [(GS1 + GS2 + GS3 + GS4)/CS+], for details see Lenaert et al., 2016). In other words, the GI indicates the reaction to all GSs relative to CS+. Since the fear-relevant discrimination training is designed to improve the discrimination of all GSs from CS+, the GI is highly suitable to reveal training-

related changes in discrimination. To prevent division by zero, all US-expectancy values were increased by 10 (i.e., the smallest step on the US-expectancy scale) and all values of SCR were increased by $\log_{10}(0.02 + 1)$, defined as smallest SCR reaction > 0 (see section 2.2.1.6). Then, ANCOVAs on GI post training were performed including fear relevance (relevant_DT, irrelevant_DT) and feedback (with, without) as between-subjects factors and GI pre training as covariate. The ANCOVA is a commonly used method for comparing pre-post change across groups, because of its good statistical power and only slight bias for floor effects (Jennings & Cribbie, 2016).

The significance level was set at $p < .05$. The Greenhouse-Geisser correction of degree of freedom was applied if the sphericity assumption was violated. For post-hoc tests, I used simple contrasts, Bonferroni corrected. Partial eta squared are indicated for effect size.

2.2.2 Results

2.2.2.1 Acquisition of conditioned fear

Successful fear acquisition is indicated by significant Stimulus × Block interactions for all dependent variables, i.e., US-expectancy ($F(1,79) = 19.41$, $p < .001$, $\eta_p^2 = .20$; **Figure 2** A), arousal ($F(1.89,149.12) = 86.34$, $p < .001$, $\eta_p^2 = .52$; **Figure 2** B), and valence ratings ($F(1.86,146.81) = 82.83$, $p < .001$, $\eta_p^2 = .51$; **Figure 2** C) and for SCR ($F(1.31,96.87) = 6.25$, $p = .008$, $\eta_p^2 = .08$; **Figure 2** D). Post-hoc contrasts confirmed that ratings were higher for CS+ vs. CS- regarding US expectancy (Acquisition1: $F(1,79) = 343.24$, $p < .001$, $\eta_p^2 = .81$; Acquisition2: $F(1,79) = 673.82$, $p < .001$, $\eta_p^2 = .90$; Bonferroni corrected $\alpha < .025$), arousal (Acquisition1: $F(1,79) = 98.12$, $p < .001$, $\eta_p^2 = .55$; Acquisition2: $F(1,79) = 165.31$, $p < .001$, $\eta_p^2 = .68$; Bonferroni corrected $\alpha < .017$), and valence (Acquisition1: $F(1,79) = 59.85$, $p < .001$, $\eta_p^2 = .431$

Acquisition2: $F(1,79) = 129.52$, $p < .001$, $\eta_p^2 = .621$; Bonferroni corrected $\alpha < .017$), while no CS+ vs. CS- differences were evident for the habituation phase (arousal: $F(1,79) = 0.62$, $p = .434$, $\eta_p^2 = .01$; valence: $F(1,79) = 0.05$, $p = .829$, $\eta_p^2 < .01$). Post-hoc contrasts for SCR also revealed no CS+ vs. CS- differences for habituation phase ($F(1,74) = 2.24$, $p = .139$, $\eta_p^2 < .03$). Regarding acquisition phase, physiological arousal was higher to CS+ vs. CS- for Acquisition 1 ($F(1,74) = 11.81$, $p < .001$, $\eta_p^2 = .14$), but not for Acquisition 2 ($F(1,74) = 1.50$, $p = .225$, $\eta_p^2 = .02$), possibly due to habituation effects.

These ANOVAs also returned significant main effects of stimulus for all dependent variables (US expectancy: $F(1,79) = 538.75$, $p < .001$, $\eta_p^2 = .87$; arousal: $F(1,79) = 124.05$, $p < .001$, $\eta_p^2 = .61$; valence: $F(1,79) = 73.29$, $p < .001$, $\eta_p^2 = .48$; SCR: $F(1,72) = 7.12$, $p = .009$, $\eta_p^2 = .09$), and main effects of block for arousal ($F(1.54,121.81) = 45.51$, $p < .001$, $\eta_p^2 = .377$), valence ($F(1.63,129.00) = 14.08$, $p < .001$, $\eta_p^2 = .15$), and SCR ($F(1,72) = 24.44$, $p < .001$, $\eta_p^2 = .06$), but not for US expectancy ($F(1,79) = 0.01$, $p = .907$, $\eta_p^2 < .01$).

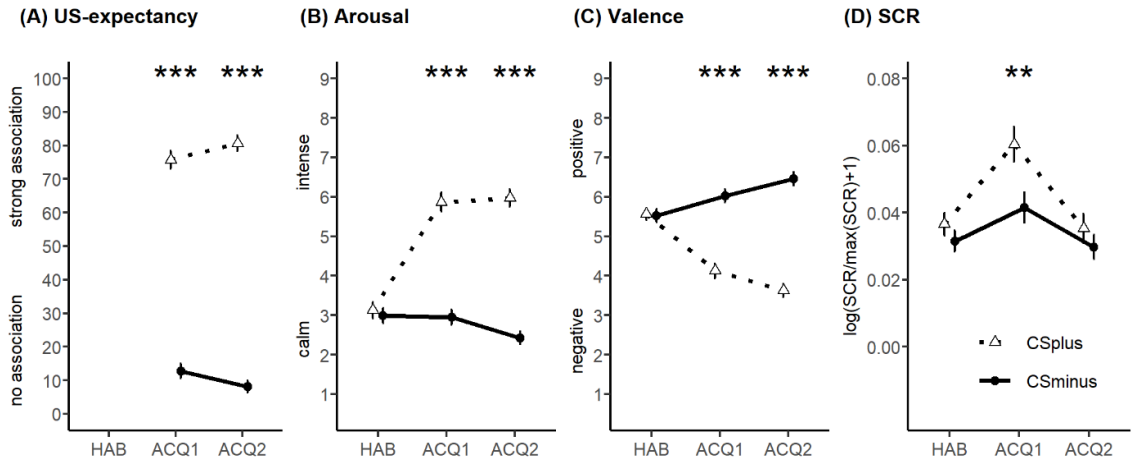


Figure 2. Habituation and acquisition of Study 1A.

Notes. Means (with SEs) of US-expectancy (A), arousal (B), and valence (C) ratings and skin conductance responses (SCR; D) are shown for CS+ (white) and CS- (black) for habituation (HAB) and for the acquisition blocks (ACQ1, ACQ2). Significance symbols indicate simple post-hoc contrasts, *** p < .001, ** p < .01.

2.2.2.2 Discrimination training effects

For US-expectancy ratings (the corresponding generalization gradients are depicted in **Figure 3** A-D), the ANCOVA on GIs revealed significant main effects of training ($F(1, 75) = 6.56$, $p = .012$, $\eta_p^2 = .08$; **Figure 4** A) and feedback ($F(1, 75) = 7.27$, $p = .009$, $\eta_p^2 = .09$; **Figure 4** B), but not their interaction ($F(1, 75) = 0.04$, $p = .850$, $\eta_p^2 < .01$; **Figure 3** E), indicating that fear generalization was reduced more effectively by fear-relevant or rewarding feedback conditions.

For arousal and valence ratings (for means and *SD*s see **Table 2**, the corresponding generalization gradients are depicted in **Figure 5** A-B), the ANCOVAs on GIs revealed no effects involving the between-factors (all *p* values > .106).

The main effect of covariate was significant for all ratings (US expectancy: $F(1, 75) = 72.13$, $p < .001$, $\eta_p^2 = .49$; arousal: $F(1, 75) = 9.71$, $p = .003$, $\eta_p^2 = .11$; valence: $F(1, 75) = 14.92$, $p < .001$, $\eta_p^2 = .17$) suggesting positive associations between

generalization assessed pre and post training (US expectancy: $r(79) = 0.69$, $p < .001$; arousal: $r(79) = 0.35$, $p = .001$; valence: $r(79) = 0.72$, $p < .001$).

Analysis of SCR with ANCOVA on GIs (see **Table 2**, the corresponding generalization gradients are depicted in **Figure 5** C) revealed a marginal significant interaction of fear relevance and feedback ($F(1, 70) = 3.7$, $p = .059$, $\eta_p^2 = .05$) only (all other p values > .153). Bonferroni corrected ($\alpha < .012$) post-hoc simple contrasts did not confirm significant impact of the combination of fear relevance and feedback (all p values > .043).

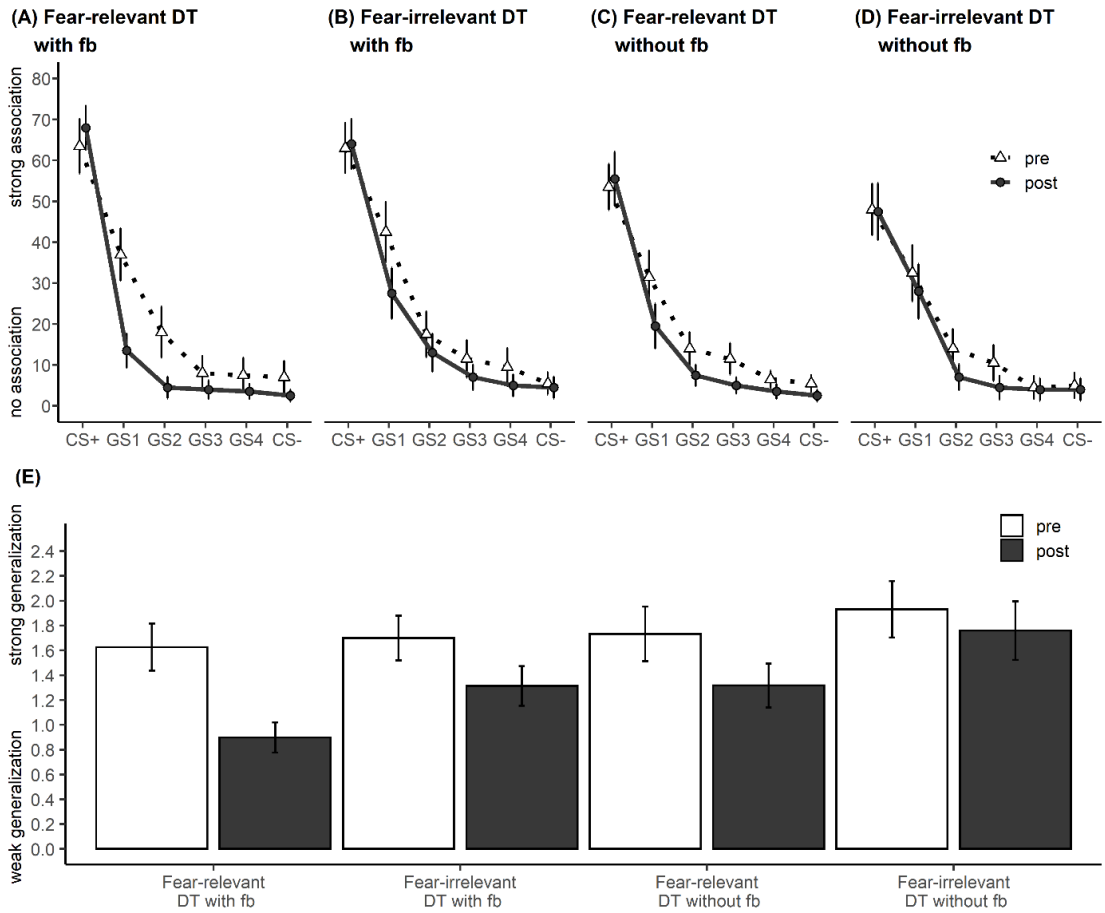


Figure 3. Generalization of US expectancy pre and post training.

Notes. Generalization gradients (A - D) and Generalization Indices (GI; E) are depicted separately for each training group. *DT* discrimination training, *fb* feedback, depicted are Means with SEs.

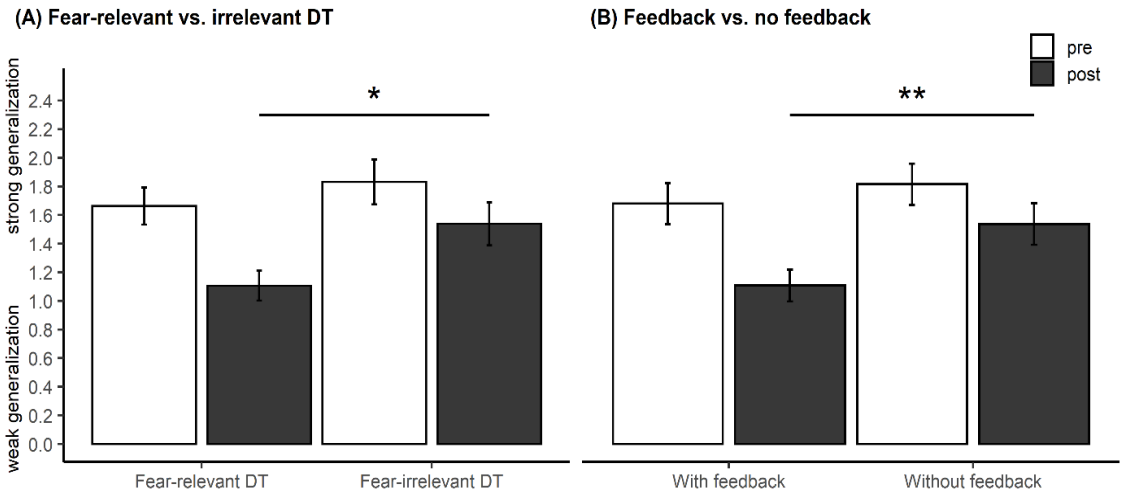


Figure 4. Generalization of US expectancy averaged by fear relevance or feedback.

Notes. Bar graphs show generalization indices (GI) with means and SEs averaged by fear relevance (A) or feedback (B) pre and post training, respectively. Significance symbols indicate main effects of ANCOVA, * p < .05., ** p < .01.

Table 2 *Generalization Indices (GI) for arousal and valence ratings as well as SCR, separately for each training group of Study 1A*

	relevant_DT _FB	*irrelevant_DT _FB*	*relevant_DT _noFB*	*irrelevant_DT _noFB*
	Arousal			
pre (*SD*)	2.60 (0.70)	2.95 (1.09)	2.15 (0.83)	2.69 (0.76)
post (*SD*)	2.11 (1.04)	2.51 (1.45)	2.05 (0.82)	3.26 (2.55)
	Valence			
pre (*SD*)	2.81 (0.87)	3.41 (2.64)	2.98 (0.74)	3.23 (0.94)
post (*SD*)	2.58 (1.01)	2.71 (1.04)	2.62 (0.79)	3.11 (1.11)
	SCR			
pre (*SD*)	3.78 (2.48)	4.15 (3.05)	4.23 (3.01)	5.34 (4.28)
post (*SD*)	5.80 (5.71)	9.83 (10.52)	7.63 (6.90)	5.41 (5.23)

Notes. DT discrimination training, *FB* feedback.

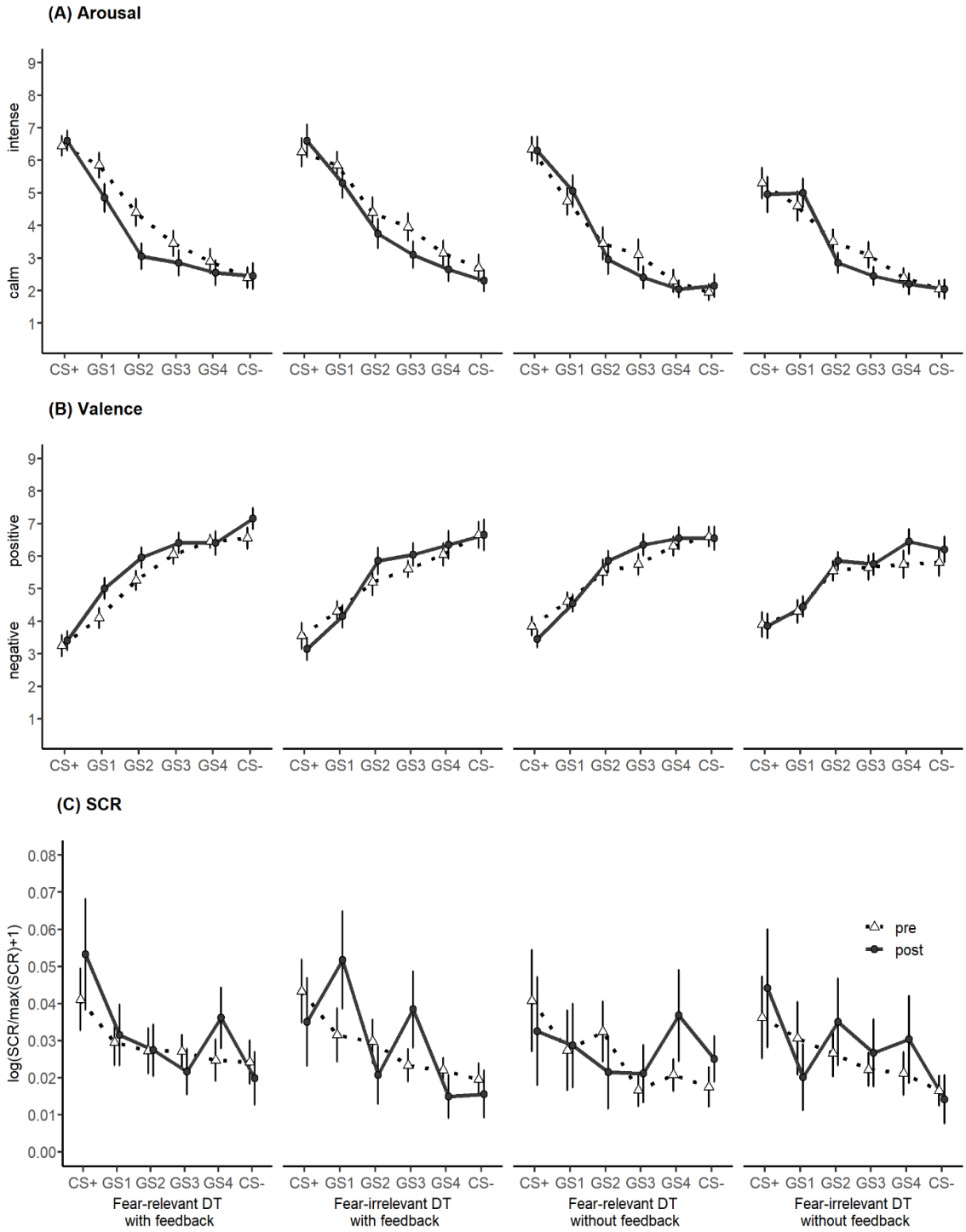


Figure 5. Generalization gradients divided by training group.

Notes. Panels show gradients for arousal (A) and valence ratings (B) as well as SCR (C) separately for generalization pre and post training. *DT* discrimination training, plotted are Means with SEs.

2.2.3 Discussion

In Study 1A, I aimed to test the hypothesis that generalization of conditioned fear to perceptually similar stimuli would be significantly better reduced by discrimination training with fear-relevant vs. fear-irrelevant stimuli. In addition, I hypothesized that discrimination-training effects would benefit from reinforcing feedback, with the largest effects resulting from fear-relevant training with feedback.

Analysis of acquisition data confirmed successful fear conditioning, as the threat cue (CS+) elicited greater arousal, more negative valence, greater US expectancy, and greater SCRs compared with the safety cue (CS-). Analysis of CS+ and CS- responses during generalization testing before and after discrimination training (see Appendix 6.1.3) shows that these conditioned fear responses remained stable throughout the experiment. Overall, these results are in accordance with previous studies with similar paradigms (Lissek et al., 2008; Schiele et al., 2016) and suggest a successful acquisition of fear towards the CS+.

Analyses of discrimination-training effects on fear generalization revealed stronger effects for the fear-relevant vs. the fear-irrelevant discrimination training, suggesting that a reduction of fear generalization is especially effective when the discrimination training is carried out with fear-relevant stimuli vs. fear-irrelevant stimuli. This result has been observed for US-expectancy ratings and is consistent with previous reports that discriminative training with fear-relevant stimuli is able to reduce fear generalization (Lommen et al., 2017), although in this study, participants were trained before fear acquisition by analogy with prevention, whereas I trained participants after fear acquisition by analogy with therapy. Moreover, it is consistent with more general findings on the transfer of discrimination learning to a test phase, which was found to be more effective after training relevant stimuli compared to

irrelevant stimuli (Furmanski & Engel, 2000). In the present study, both the fear-relevant and fear-irrelevant training groups improved discrimination performance during training (see Appendix 6.1.2 for further details). Consequently, the more efficient fear reduction of the fear-relevant training group might be related to a more effective transfer to the generalization test after training. However, we know from previous studies that discrimination training, even with fear-irrelevant stimuli, generally reduces subsequent generalization effects (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017). This may explain why fear relevance effects are present but rather small. Secondly, analyses of discrimination-training effects on fear generalization revealed stronger effects for discrimination training with reinforcing feedback vs. without feedback. Again, the effect was observed for US-expectancy ratings. From this, I conclude that a reduction in fear generalization is particularly effective when discrimination training is conducted with reinforcing feedback compared to no feedback. However, no support was found for my hypothesis that the combination of training, i.e., training with fear-relevant stimuli *and* feedback, is especially effective in reducing fear generalization. One explanation could be that feedback affects very general processes that are independent of the processing and discrimination of the presented stimuli. Previous studies have shown that feedback improves performance in general (Sasaki et al., 2010; Weil et al., 2010) as well as mood (Westermann, Spies, Stahl, & Hesse, 1996). Moreover, reward facilitates attentional processes towards the reward-associated cue (Chelazzi, Perlato, Santandrea, & Della Libera, 2013) and increases motivation (Fishbach, Eyal, & Finkelstein, 2010). It must be acknowledged that due to a programming error, data on discrimination training performance is not available for the groups without feedback. Therefore, I cannot test whether and how feedback improved discrimination performance during training, and consequently, I cannot relate the effects of feedback on performance to subsequent generalization

effects. Nonetheless, it is likely that the implemented reinforcing feedback improved the participants' mood, attention, and motivation in general, which consequently improved discrimination performance during the subsequent generalization test.

As noted earlier, the effects of the applied discrimination training were limited to US-expectancy ratings. One possible explanation lies in the typology of the ratings. As proposed previously (Lonsdorf et al., 2017), US expectancy reflects cognitive learning processes, whereas valence and arousal ratings reflect more affective learning processes. Therefore, new learning experiences are found more easily and earlier in US-expectancy ratings. Correspondingly, the training conducted here positively affected US-expectancy ratings. In contrast, extinction following evaluative conditioning, which is reflected in changes in valence, is difficult to achieve (W. Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006). This might also be true for discrimination learning, which could explain why I did not find training effects for affective ratings. Moreover, the dissociation between affective and cognitive ratings suggests that other (higher-order) processes (e.g., WM and inference rules, for a more detailed discussion see sections 2.4 and 4.2) besides perception might have determined fear generalization and its reduction in the present paradigm.

By failing to confirm any effects of fear relevance or feedback on physiological arousal, i.e., SCRs, the effectiveness of this "therapeutic" approach of discrimination training is limited to the verbal level. In contrast to all ratings, SCRs post training were not significantly affected by the pre-training level, as shown by the absent main effect of covariate. This means that all trainings affected the electrodermal activity to such an extent that the generalization pattern pre training was resolved. Presumably, the main reason stems from the nature of the training. While participants passively observed the

stimuli during the generalization blocks, they actively performed a task during the discrimination training. The active task might have superimposed specific training effects. In addition, the discriminative physiological arousal established during the first acquisition block disappeared during the second acquisition block, which is likely due to habituation and could explain the lack of training effects for SCR.

In summary, Study 1A provided first evidence for more successful reduction of already occurred fear generalization using a fear-relevant vs. fear-irrelevant discrimination training, at least for cognitive fear parameters, i.e., US expectancy. Furthermore, the study showed that reinforcing feedback during discrimination training reduced generalization of US expectancy, possibly because feedback generally increases attention, motivation, and mood.

2.3 Study 1B: Fear-reducing effects of discrimination training compared to non-discriminative control training

The results of Study 1A showed that a reduction of fear generalization is especially effective when the discrimination training is carried out with fear-relevant stimuli vs. fear-irrelevant stimuli or with reinforcing feedback vs. no feedback. However, the superiority of the training with fear-relevant stimuli was rather small, probably because discrimination trainings, also with fear-irrelevant stimuli, reduce subsequent generalization effects in general (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017). Thus, it would be interesting to determine the discrimination-training effects in comparison with a control condition without discrimination training. Another question that came up in Study 1A was if feedback improves mood, attention and motivation in general and thereby reduces fear generalization independently from discrimination processes. This question can also be clarified with the help of a control

condition without discrimination training but a motivation condition, e.g., a task with comparable cognitive load and reinforcing feedback for correct responses.

Therefore, to examine these questions, two further control groups were recruited in Study 1B. These groups underwent the same differential fear conditioning and generalization paradigm (Baumann et al., 2017; Schiele et al., 2016) as did the discrimination-training groups (see Study 1A). The experiment differed in the training part between generalization blocks only. Here, the control groups completed a non-discriminative training in which they had to solve a simple arithmetic problem. The task occurred either with or without feedback on arithmetic performance. I hypothesized that the discrimination training reduces fear generalization more effectively than the non-discriminative control training. Besides, based on the results of the previous study, I speculated that feedback contributes to fear reduction independently from discrimination.

2.3.1 Materials and methods

2.3.1.1 Participants

Some analyses of Study 1B included two groups already assessed in Study 1A, i.e., the fear-relevant and fear-irrelevant discrimination-training groups with feedback (**Table 3**), in order to compare them with newly collected data of the non-discriminative control-training group. For the latter, 40 further participants were invited to the laboratory. Like the participants of Study 1A, they were recruited on the Internet platform SONA (psywue.sona-systems.com) by the University of Würzburg and screened for the same exclusion criteria and age span (see section 2.2.1.1). After arrival, all participants read and signed the informed consent and then were randomly assigned to one of two new experimental groups (for details see **Table 3**). The newly assessed groups with non-discriminative training showed no pre-experimental differences

regarding gender, age, native language or handedness. Besides, there was no difference between the feedback groups with non-discriminative task and discrimination task for these sample descriptives.

For further sample description, participants filled in the German version of the STAI (Laux et al., 1981) as well as of the BDI (Hautzinger et al., 2006) before the experiment (**Table 3**). Participants of the non-discriminative training group with feedback had higher STAI trait scores than participants who completed the same task without feedback ($F(1, 38) = 6.38$, $p = .016$, $\eta_p^2 = .14$; **Table 3**). Moreover, the STAI trait score of the feedback group with non-discriminative training differed from that of feedback groups with discrimination trainings ($F(2, 57) = 3.81$, $p = .028$, $\eta_p^2 = .12$; **Table 3**) in that the non-discriminative group scored higher compared to the fear-relevant discrimination training group ($F(1, 57) = 7.55$, $p = .008$, $\eta_p^2 = .12$; all further p values > .117). Instead, no group differences were observed for BDI (**Table 3**). To exclude that training effects occurred only due to the differences in STAI trait, I repeated all analyses controlling for STAI trait by including it as covariate. Since the training effects were unchanged, results of these ANCOVAs are not reported.

Table 3 *Sample characteristics of Study 1B*

	relevant_DT_FB	*irrelevant_DT_FB*	*non-DT_FB*	*non-DT_noFB*	*comparisons trainings*	*comparisons non-DT groups*
N	20	20	20	20		
gender (♀)	10	10	9	10	$\chi(2) = 0.13$, $p = .935$	$\chi(1) = 0$, $p = 1$
age (*SD*)	26.10 (8.42)	24.75 (5.37)	27.35 (7.90)	24.20 (4.60)	$F(2,57) = 0.63$, $p = .538$	$F(1,38) = 2.08$, $p = .157$
language (german)	17	19	18	18	$\chi(2) = 0.44$, $p = .804$	
handedness (right)	19	17	20	20	$\chi(2) = 3.75$, $p = .153$	
STAI (*SD*)	32.70 (6.53)	35.25 (7.41)	38.75 (6.92)	33.70 (5.67)	$F(2,57) = 3.81$, $p = .028$	$F(1,38) = 6.38$, $p = .016$
BDI (*SD*)	3.70 (4.46)	5.55 (6.35)	5.75 (4.35)	6.25 (8.40)	$F(2,57) = 0.97$, $p = .386$	$F(1,38) = 0.06$, $p = .814$

Notes. DT discrimination training, *FB* feedback, *STAI* State-Trait-Anxiety-Inventory Trait Scale, *BDI* Beck Depression Inventory II.

2.3.1.2 Questionnaires

Same as in Study 1A, new participants were asked to provide some demographic information, e.g., age, gender, native language, and handedness, and completed the German versions of the BDI, STAI trait, STAI state, and PANAS (see section 2.2.1.2 for description of the questionnaires; see Appendix 6.2.1 for analyses of state questionnaires).

2.3.1.3 Stimulus material

Stimulus material of the conditioning and generalization experiment were identical with Study 1A (for details see section 2.2.1.3). In the non-discriminative control training, a new, neutral stimulus, i.e., a blue square (a polygon graphic with

radius 80, which corresponds with a side length of about 113, created with Presentation software Version 16.0; Neurobehavioral Systems, Inc., Albany, CA), was introduced (also see section 2.3.1.5).

2.3.1.4 Procedure

The experimental procedure (**Figure 1** A) of the newly assessed groups differed from Study 1A only in one point, the non-discriminative control training (**Figure 6**), which is explained in more detail in the following paragraph.

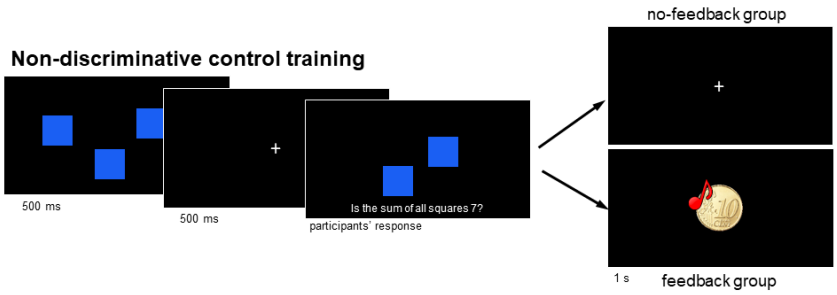


Figure 6. Overview of the non-discriminative control training.

Notes. Control training was embedded in two generalization blocks and completed either with or without feedback on performance.

2.3.1.5 Non-discriminative control training

The protocol of the control training (**Figure 6**) was based very closely on the protocol of the discrimination trainings described in Study 1A. Instead of the faces or lines of the discrimination trainings, square patterns were presented. One square pattern consisted of two to five blue squares that were displayed on the screen. Every trial consisted of three elements: a 500 ms square pattern presentation, followed by a 500 ms

inter-stimulus interval (ISI), and another square pattern presentation. Below the second square pattern the question “Is the sum of all squares 7?” was displayed. Participants were asked to solve this simple arithmetic task and respond by pressing the S and L on a German keyboard, which were assigned to the answers yes or no respectively. Response time was not limited. Therefore, square pattern and question lasted on the screen until response was given.

In one group of participants (non-DT_FB), correct answers were reinforced by feedback, consisting of a 10 Euro cent picture and a cash register sound (52.4 dB) lasting 1 s. Incorrect answers were not followed by feedback. In another group (non-DT_noFB), no feedback was provided at any time. Altogether, participants performed 50 training trials with a short break after 25 trials. In all trainings, trials were separated by an ITI lasting 1-3 s randomly. For 20 out of the 50 trials the total sum of squares was seven (two – five, three – four, four – three, five – two, 5 times each). For the remaining 30 trials, the number of squares differed from seven, with more trials having a sum close to seven (2 x two – two, 3 x two – three, 3 x three – two, 2 x two – four, 2 x three – three, 4 x four – two, 2 x three – five, 2 x four – four, 3 x five – three, 2 x four – five, 3 x five – four, 2 x five – five). Accordingly, in contrast to the discrimination trainings, in the non-discriminative control training, the first stimulus of each trial was not always the same; otherwise, the arithmetic task would have been too simple. The arrangement of squares within each pattern was not random. Instead, there were two variants of square patterns with two to five squares, respectively.

2.3.1.6 Data recording and reduction

Recording and reduction of the newly collected data was done in the same way as for Study 1A (see section 2.2.1.6). Participants with an overall raw mean response smaller 0.02 µS were considered non-responders and therefore excluded from statistical

testing of SCR. Accordingly, *n* = 20 of the relevant_DT_FB group, *n* = 16 of the relevant_DT_noFB group, *n* = 18 of the non-DT_FB group and *n* = 15 of the non-DT_noFB group were included in statistical analysis of SCR.

2.3.1.7 Statistical analysis

Statistical analyses were carried out in the R software environment (version 3.6.1) using packages 'afex' (version 0.26-0; Singmann et al., 2015) and 'emmeans' (version 1.4.5; Lenth & Lenth, 2018).

Effects of fear acquisition were analyzed with ANOVAs. They presented the within-subject factors stimulus (CS+, CS-) and block, with the factor block having three levels (habituation, acquisition 1, acquisition 2) for arousal, valence and SCR data, and two levels (acquisition 1, acquisition 2) for US-expectancy data.

In order to investigate training effects, the Generalization Index (GI = [(GS1 + GS2 + GS3 + GS4)/CS+], for details see section 2.2.1.7 and Lenaert, van de Ven, Kaas, & Vlaeyen, 2016) was calculated for each generalization block.

For analyses of discrimination-training effects in comparison with the non-discriminative control condition, ANCOVAs on GI post training were performed including training (relevant_DT_FB, irrelevant_DT_FB, non-DT_FB) as between-subjects factor and GI pre training as covariate. Additionally, I analyzed the groups' performance during training with an ANOVA having part (part 1, part 2) as within-subject factor and training (relevant_DT_FB, irrelevant_DT_FB, non-DT_FB*)* as between-subjects factor.

To analyze whether feedback has a general effect independently from discrimination, the non-discriminative control training groups were compared with

ANCOVAs on GI post training having feedback (with, without) as between-subjects factor and GI pre training as covariate.

The significance level was set at $p < .05$ and, if necessary, the Greenhouse-Geisser correction of degree of freedom was applied. For post-hoc tests, I used simple contrasts, Bonferroni corrected. Effect sizes are reported as partial eta squared.

2.3.2 Results

2.3.2.1 Acquisition of conditioned fear in non-discriminative training groups

Fear acquisition was successful for US-expectancy ratings (**Figure 7** A) as indicated by a significant main effect of stimulus ($F(1, 39) = 247.69$, $p < .001$, $\eta_p^2 = .86$) as well as its interaction with block ($F(1, 39) = 15.32$, $p < .001$, $\eta_p^2 = .28$, all further p values > .127. Post-hoc contrasts confirmed stronger association of the US with CS+ compared to CS- in both blocks (Acquisition1: $F(1, 39) = 160.07$, $p < .001$, $\eta_p^2 = .80$; Acquisition2: $F(1, 39) = 268.09$, $p < .001$, $\eta_p^2 = .87$; Bonferroni corrected $\alpha < .025$). Successful fear acquisition is also reflected in arousal (**Figure 7** B) and valence (**Figure 7** C) ratings. ANOVAs returned significant main effects of stimulus (arousal: $F(1, 39) = 45.12$, $p < .001$, $\eta_p^2 = .54$, valence: $F(1, 39) = 25.70$, $p < .001$, $\eta_p^2 = .40$) and block (arousal: $F(2, 78) = 13.69$, $p < .001$, $\eta_p^2 = .26$, valence: $F(2, 78) = 7.61$, $p < .001$, $\eta_p^2 = .16$) as well as their interaction (arousal: $F(1.55, 60.31) = 25.38$, $p < .001$, $\eta_p^2 = .39$, valence: $F(1.69, 65.95) = 38.27$, $p < .001$, $\eta_p^2 = .50$). Post-hoc contrasts confirmed that ratings were higher for CS+ vs. CS- regarding arousal (Acquisition1: $F(1, 39) = 44.08$, $p < .001$, $\eta_p^2 = .53$; Acquisition2: $F(1, 39) = 55.76$, $p < .001$, $\eta_p^2 = .59$; Bonferroni corrected $\alpha < .017$), and valence (Acquisition1: $F(1, 39) = 25.9$, $p < .001$, $\eta_p^2 = .40$; Acquisition2: $F(1, 39) = 41.87$, $p < .001$, $\eta_p^2 = .52$; Bonferroni corrected $\alpha < .017$) in

both acquisition blocks, while no CS+ vs. CS- differences were evident for the habituation block (arousal: $F(1, 39) = 0.22$, $p = .639$, $\eta_p^2 < .01$; valence: $F(1, 39) = 0.04$, $p = .852$, $\eta_p^2 < .01$). For SCR, the ANOVA revealed a significant main effect of stimulus ($F(1, 32) = 4.23$, $p = .048$, $\eta_p^2 = .12$) indicating stronger reactions to CS+ vs. CS- during all blocks and a main effect of block ($F(1.59, 50.72) = 11.03$, $p < .001$, $\eta_p^2 = .26$), but no Stimulus × Block interaction ($F(1.29, 41.29) = 1.05$, $p = .331$, $\eta_p^2 = .03$; **Figure 7** D). Post-hoc contrasts of Block showed stronger SCRs to both stimuli during Acquisition1 compared to habituation ($F(1, 32) = 11.99$, $p = .002$, $\eta_p^2 = .27$) or Acquisition2 ($F(1, 32) = 15$, $p < .001$, $\eta_p^2 = .32$), while response strength did not differ between habituation and Acquisition2 ($F(1, 32) = 0$, $p = .978$, $\eta_p^2 < .01$).

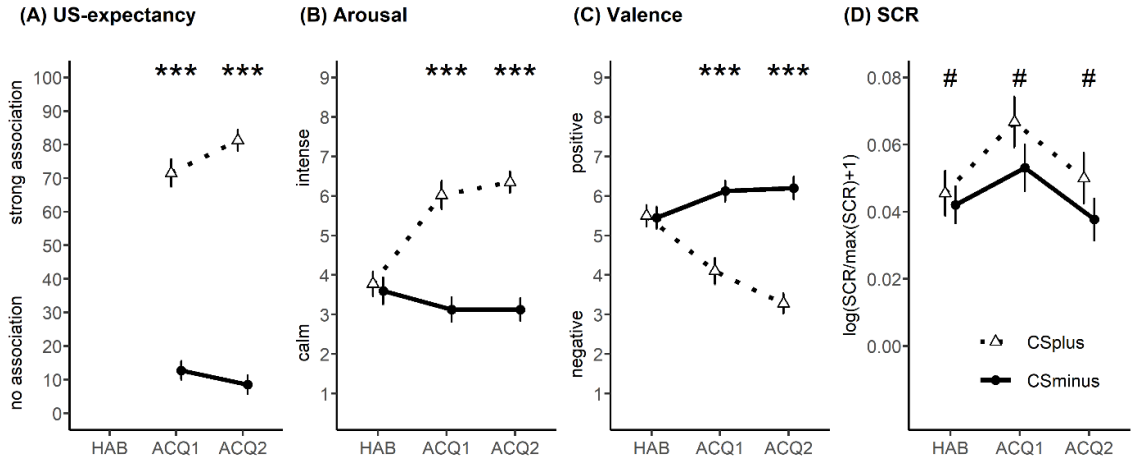


Figure 7. Habituation and acquisition of the non-discriminative control training groups.

Notes. Plots show ratings of US expectancy (A), arousal (B), and valence (C), and skin conductance responses (D). Mean values (with SEs) of ratings and SCRs are shown for CS+ (white) and CS- (black) for habituation (HAB) and for the acquisition blocks (ACQ1, ACQ2). Significance symbols indicate simple post-hoc contrasts, *** p < .001, # p < .10.

2.3.2.2 Discrimination vs. non-discriminative training effects

Neither for ratings nor for SCR, the ANCOVAs on GIs revealed a significant main effect of training (all p values > .117; see **Table 4** and for US-expectancy ratings

also **Figure 8**), indicating that the discrimination trainings do not exceed the fear reduction observed after the non-discriminative control training.

The main effect of covariate was significant for US-expectancy ratings ($F(1, 56) = 18.19$, $p < .001$, $\eta_p^2 = .25$) and valence ratings ($F(1, 56) = 6.82$, $p = .012$, $\eta_p^2 = .11$), suggesting positive associations between generalization assessed pre and post training (US expectancy: $r(58) = .50$, $p < .001$; arousal: $r(58) = .34$, $p = .009$). No effect of covariate was observed for subjective or physiological arousal (ratings: $F(1, 56) = 1.56$, $p = .217$, $\eta_p^2 = .03$; SCR: $F(1, 53) = 2.21$, $p = .143$, $\eta_p^2 = .04$).

Table 4 *Generalization Indices (GI) separately for groups of Study 1B*

	relevant_DT _FB	*irrelevant_DT _FB*	*non-DT _FB*	*non-DT _noFB*
	US expectancy			
pre (*SD*)	1.63 (0.85)	1.73 (0.99)	2.03 (1.19)	1.73 (1.08)
post (*SD*)	0.90 (0.54)	1.32 (0.79)	1.28 (0.91)	1.32 (0.85)
	Arousal			
pre (*SD*)	2.60 (0.70)	2.95 (1.09)	4.18 (4.40)	2.59 (0.84)
post (*SD*)	2.11 (1.04)	2.51 (1.45)	2.55 (1.24)	3.46 (5.16)
	Valence			
pre (*SD*)	2.81 (0.87)	3.41 (2.64)	2.95 (0.78)	3.11 (1.05)
post (*SD*)	2.58 (1.01)	2.71 (1.04)	2.54 (0.85)	2.52 (0.90)
	SCR			
pre (*SD*)	3.78 (2.48)	4.15 (3.05)	4.35 (2.96)	3.77 (2.21)
post (*SD*)	5.80 (5.71)	9.83 (10.52)	4.96 (5.61)	5.04 (6.53)

Notes. DT discrimination training, *FB* feedback.

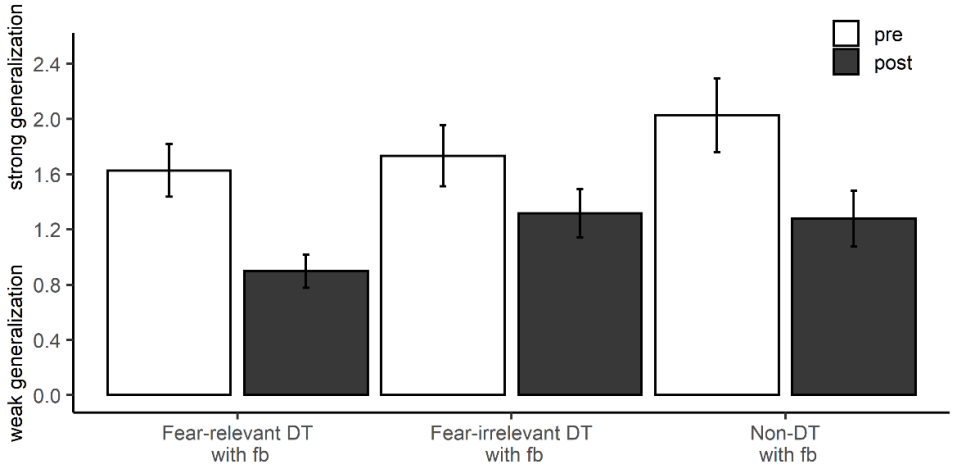


Figure 8. Generalization of US expectancy divided by training condition.

Notes. DT discrimination training*, Non-DT* non-discriminative control training*, fb* feedback*, d*epicted are Generalization Indices (GI) with Means and SEs separately for generalization pre and post training.

2.3.2.3 General feedback effect

Generalization gradients of the non-discriminative training groups are depicted in **Figure 9** A-B and **Figure 10**. For none of the outcome measures, the ANCOVAs on GIs revealed a significant main effect of feedback (all *p* values > .376; see **Table 4** and for US-expectancy ratings also **Figure 9** C), indicating that feedback during the non-discriminative control training does not facilitate reduction of fear in the subsequent generalization test.

The main effect of covariate was significant for all ratings (US expectancy: $F(1, 37) = 5.81$, $p = .021$, $\eta_p^2 = .14$; valence: $F(1, 37) = 5.67$, $p = .023$, $\eta_p^2 = .13$) except arousal ($F(1, 56) = 1.56$, $p = .217$, $\eta_p^2 = .03$). The present effects suggest positive associations between generalization assessed pre and post training (US expectancy: $r(38) = .36$, $p = .022$; arousal: $r(38) = .36$, $p = .022$). For SCR, the main effect of covariate resulted marginal significant ($F(1, 53) = 2.21$, $p = .143$, $\eta_p^2 = .04$), but was not confirmed by significant correlation of the generalization assessed pre and post training ($r(37) = .26$, $p = .105$).

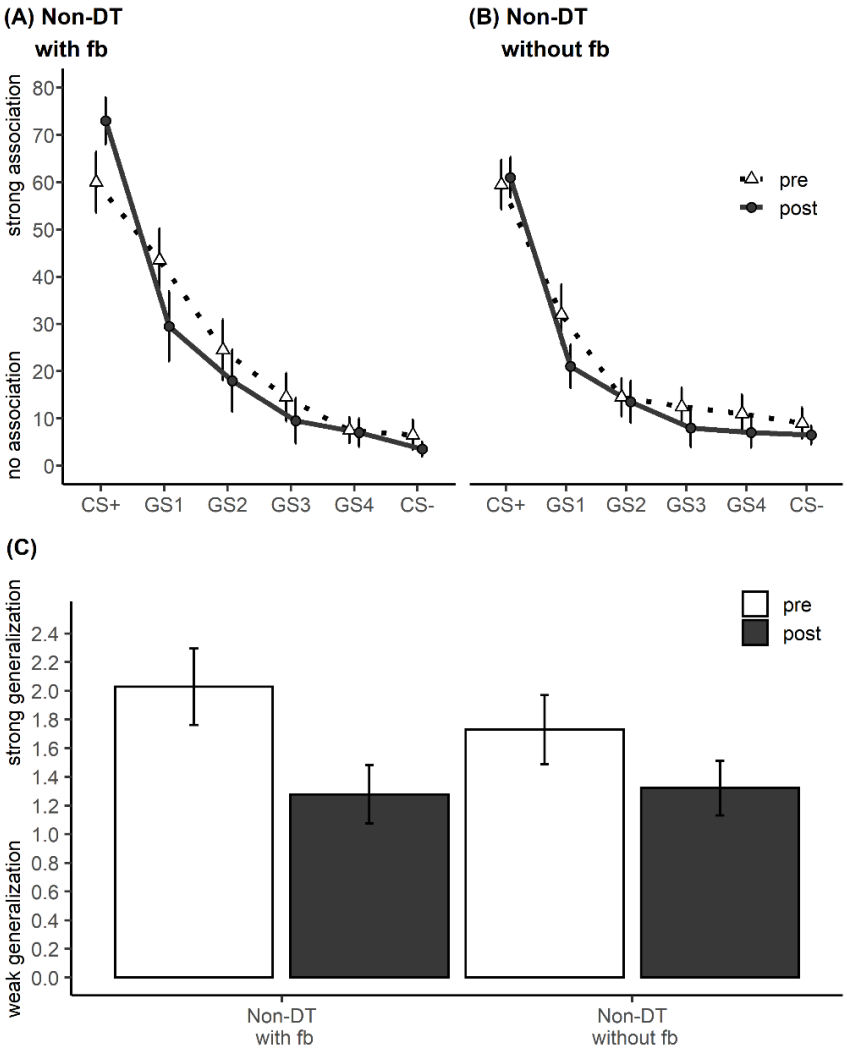


Figure 9. Generalization of US expectancy in non-discriminative control training groups with and without feedback.

Notes. Plotted are Means with SEs of generalization gradients (A - B) and Generalization Indices (GI; C) pre and post training, respectively, *Non-DT* non-discriminative control training groups, *fb* feedback.

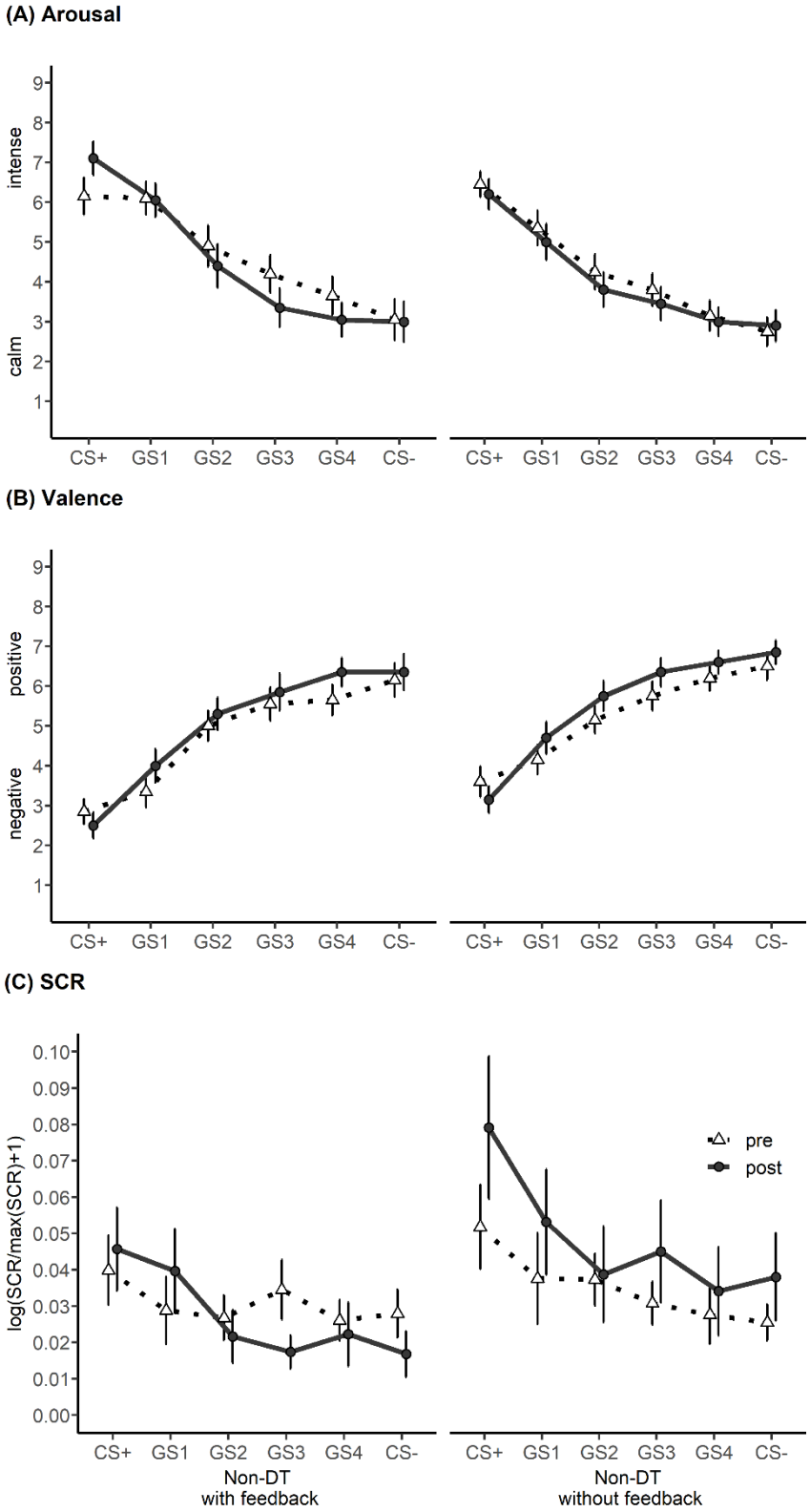


Figure 10. Generalization gradients of non-discriminative control training groups with or without feedback.

Notes. Panels show gradients (with Means with SEs) for arousal (A) and valence ratings (B) as well as SCR (C) separately for generalization pre and post training. *Non-DT* non-discriminative control training.

2.3.3 Discussion

The goal of Study 1B was to show that both fear-relevant and fear-irrelevant discrimination trainings reduce fear generalization more effectively than a non-discriminative control training. Furthermore, I tested the hypothesis that feedback contributes to fear reduction independently from discrimination.

As for the re-used groups of Study 1A, fear conditioning was also obtained in the newly assessed groups, which later on underwent a non-discriminative control training, as illustrated by higher responses to the threat cue (CS+) vs. the safety cue (CS-) across all outcome measures. These results show that the paradigm is well-established (Haddad et al., 2012; K. Herzog et al., 2021; Lau et al., 2008; Schiele et al., 2016; Wurst et al., 2021).

Against my expectation, the non-discriminative control training turned out to be no less effective in reducing fear generalization pre to post training compared to both fear-relevant and fear-irrelevant discrimination trainings. According to the analysis, all three groups showed comparable levels of fear generalization post training when controlling for generalization pre training. This result is in contradiction to that of others, whose discrimination training, compared to a control task, reduced the occurrence of fear generalization (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017; Lommen et al., 2017). A comparison of the used control conditions provides a plausible explanation. In order to ensure high comparability of discrimination training and non-discriminative control training, the two trainings of the current study were identical in terms of trial structure and time course. More precisely, stimulus presentations of one discrimination trial are separated by an ISI and the combination of both trial parts is necessary to make the perceptual judgement. Similarly, during control training trials, stimulus presentations before and after an ISI must be combined to solve

the arithmetic problem. In contrast, the control tasks developed by Ginat-Frolich et al. (2017) or Lommen et al. (2017) differ substantially from the discrimination trainings with regard to trial structure and time course. Their discrimination training also requires comparison of stimuli presented before and after an ISI. However, trials of the control task are less complex insofar as all stimuli are presented at the same time. Consequently, it stands to reason that all trainings of the present study equally involve several (higher-order) processes, which might facilitate perceptual discrimination in the subsequent test, e.g. working memory and attentional control (for a broader discussion see sections 2.4 and 4.2). Moreover, feedback was included in all three trainings. Since feedback seems to affect very general processes like mood (Westermann et al., 1996) and motivation (Fishbach et al., 2010), it might have contributed to the success of both discrimination trainings and non-discriminative control training (for a broader discussion also see section 2.4).

However, the hypothesis that feedback has fear-reducing effects independently from discrimination could not be confirmed when comparing the groups who underwent the non-discriminative control training with vs. without feedback. Descriptively, a slightly stronger reduction of fear generalization could be observed after the non-discriminative training with vs. without feedback. However, there was only a significant feedback effect if feedback was given during discrimination training (Study 1A). As already discussed in Study 1A, feedback affects several processes. For example, feedback can improve mood (Westermann et al., 1996) in general and subsequently reduce fear (Samsom & Rachman, 1989; Zbozinek, Holmes, & Craske, 2015). Other feedback mechanisms, on the other hand, seem to operate only within a specific task and do not simply transfer to another. Feedback increases attention towards the reward-associated cue (Chelazzi et al., 2013) and affects motivational processes by increasing the commitment to specific goals (Fishbach et al., 2010). In the present study,

participants of the non-discriminative training group received feedback on whether they added up several squares correctly. Remarkably, the non-discriminative training group with feedback made more mistakes than the fear-relevant discrimination-training group with feedback (see Appendix 6.2.2.1). Consequently, the non-discriminative training group received positive feedback in less trials, possibly affecting unfavorably the induction of positive mood. Moreover, neither the increased attention to square patterns, nor the increased motivation to improve arithmetic performance, can be used directly to improve stimulus discrimination and thereby reduce fear generalization in the subsequent generalization test. In addition, analysis of performance during the control training (see Appendix 6.2.2.2) revealed that the groups with and without feedback improved equally well. Apparently, feedback was not necessary to improve in the non-discriminative task, probably because participants were good at such easy arithmetic task already before training and only needed a short familiarization phase (Liu, Lu, & Dosher, 2010). Conversely, in the discrimination trainings participants had difficulty to distinguish the two most similar stimuli (see Appendix 6.2.2.1) so that feedback provided additional information, which was needed to improve performance (Liu et al., 2010).

In summary, both discriminative and non-discriminative trainings seem to involve several (higher-order) processes, which facilitate stimulus discrimination and consequently reduce fear generalization in the subsequent test. Moreover, feedback had fear-reducing effects only when applied during discrimination training, presumably because it specifically increases commitment to perceptual tasks. Consequently, feedback was ineffective when applied during non-discriminative control training.

2.4 Discussion Study 1

The present study was aimed at reducing generalization of conditioned fear to perceptually similar stimuli by discrimination training after the generalization has already manifested. At this point, patients are usually seeking therapy. Therefore, in order to resemble a therapeutic approach training in the present study took place after fear acquisition and between two generalization tests. Consequently, this is the first study that is capable of verifying training effects by comparing generalization indices pre vs. post training.

Different training conditions were compared. On the one hand, I expected a more effective fear reduction after discrimination training with fear-relevant vs. fear-irrelevant stimuli. Both discrimination trainings were expected to reduce generalization better than a non-discriminative control training. On the other hand, I examined whether feedback vs. no feedback on discrimination performance increases training effectiveness or whether even feedback on any task can have this effect.

Indeed, Study 1A demonstrated a somewhat more effective reduction in fear generalization when discrimination training was conducted with fear-relevant stimuli compared to fear-irrelevant stimuli, at least for US-expectancy ratings. As discussed in section 2.2.3, discrimination learning with fear-relevant vs. fear-irrelevant stimuli can be more easily transferred to a subsequent test (Furmanski & Engel, 2000). The nevertheless small effect of fear relevance might be explained by the fact that all participants were trained in discrimination. This conclusion is also supported by previous studies showing that discrimination training with fear-irrelevant stimuli also generally reduces subsequent generalization effects (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017). On this basis, both discrimination trainings should allow greater fear reduction compared to a non-discriminative control training.

However, the results of Study 1B did not confirm this hypothesis, leading to the conclusion that other (higher-order) processes besides perceptual discrimination might be involved in both discrimination and control training that contribute to fear reduction (also see section 4.2).

For example, the effect could be related to working memory (WM) processes, which are addressed in all trainings, as participants have to keep in mind the first stimulus presentation and combine it with the subsequent one to accomplish the task (Lenaert et al., 2016). The WM's ability to filter irrelevant from relevant information (Baddeley, 2012; Derakshan & Eysenck, 2009; Eysenck, Derakshan, Santos, & Calvo, 2007) may have remained active during subsequent testing and caused a reduction in fear generalization. Since the fear-relevant discrimination training precisely improves discrimination between relevant and irrelevant CS+ information, the facilitatory effect of WM on fear reduction might be particularly strong here, consistent with the effect of fear relevance in Study 1A. Nonetheless, WM might facilitate perceptual discrimination (Amitay, Zhang, Jones, & Moore, 2014) through its function in attentional control (Derakshan & Eysenck, 2009; Eysenck et al., 2007) after the other trainings as well. Perceptual discrimination post training could be further enhanced by inductive rule learning during training, e.g., that attention to visual features of stimuli is a useful strategy (Livesey & McLaren, 2009). The derivation of this rule from WM is particularly suggested in discrimination trainings and might be boosted by the fixed sequence of stimulus presentations during these trainings (always beginning with CS+ or Line1), and therefore perceptual learning might be overestimated (Garcia-Perez & Alcala-Quintana, 2020). Albeit less clearly, deriving the same rule also seems to make sense in the non-discriminative control training, as one needs to pay attention to small differences in the number of squares that the stimulus patterns contain in both parts of the trial.

One mechanism that was specifically used or omitted in both trainings was feedback. Study 1A provided evidence of a particularly effective reduction in fear generalization when discrimination training was conducted with reinforcing feedback compared to no feedback, presumably because feedback affects very general processes that are unrelated to processing and discriminating the presented stimuli. Accordingly, feedback in any task could have a fear-reducing effect. This hypothesis was tested in Study 1B but could not be confirmed. Even though feedback affects general processes such as mood (Westermann et al., 1996), attention (Chelazzi et al., 2013) and motivation (Fishbach et al., 2010), improvements in at least the latter two processes appear to act predominantly in a task-specific manner. That is, if feedback increases attention and motivation for one task (e.g., arithmetic operations), attention and motivation for another task (e.g., perceptual discrimination) may remain unaffected. In Study 1A, participants received feedback on their perceptual performance, which significantly contributed to improved perceptual discrimination also during the subsequent generalization test.

A few limitations should be considered. First, as is evident from the discussion, there are several mechanisms underlying fear generalization and perceptual learning in the present paradigm. It is difficult to separate which mechanisms were addressed during training or subsequent testing and led to fear reduction. Therefore, further research is needed to disentangle the various mechanisms that contribute to fear generalization and its reduction, e.g., through stimulus discrimination training.

Second, the effects of the conducted discrimination trainings were restricted to US-expectancy ratings and no effects were demonstrated for physiological arousal, i.e., SCRs, limiting the effectiveness of my "therapeutic" approach to discrimination training to the cognitive-verbal level. However, previous studies have not included affective

ratings (Ginat-Frolich et al., 2017; Lommen et al., 2017) or physiological measures (Lommen et al., 2017) or also have not found training effects for physiological measures (Ginat-Frolich et al., 2017).

Third, since the presented study is a proof-of-principle study, the group sizes are relatively small. Nevertheless, first conclusions can be drawn about the efficacy of a "therapeutic" approach of fear reduction by discrimination trainings, which was studied here for the first time.

Forth, I examined healthy participants, who show moderate fear generalization as also reported in previous studies (Lissek et al., 2008). Conversely, fear generalization is much more pronounced in high anxious individuals and anxiety patients, culminating in overgeneralization (Laufer & Paz, 2012; Lissek & Grillon, 2010; Lissek, Kaczkurkin, et al., 2014; Sep et al., 2019). Therefore, in future studies, the discrimination training should be applied to more clinically relevant samples.

In summary, this proof-of-principle study provided evidence for a more successful reduction of ascertained fear generalization by fear-relevant vs. fear-irrelevant discrimination training in healthy individuals, at least for cognitive fear parameters, i.e., US expectancy. Moreover, the study revealed that reinforcing feedback during discrimination training reduces generalization of US expectancy particularly well, presumably via motivational mechanisms that specifically increase commitment to perceptual tasks. It should be noted, however, that in addition to perceptual discrimination and feedback, several other (higher-order) processes may have contributed to the effectiveness of the training and subsequent reduction in fear generalization. Importantly, this study includes the first attempt to use discrimination training in a treatment-like approach, i.e., after fear acquisition and a first demonstration of fear generalization. Accordingly, this is the first indication that preexisting fear

generalization can be reduced by fear-relevant discrimination training or reinforcing discrimination feedback.

3 Study 2: Particularly effective reduction of fear generalization by discrimination training with fear-relevant stimuli or feedback - replication test and influence of risk factors for anxiety disorders

The main goal of Study 2 was to address questions I have not been able to answer in Study 1, mainly due to limitations. Essentially, three questions were addressed. First, do the effects of discrimination training from Study 1A hold in a larger sample? Second, how do individuals at risk of anxiety disorders generalize fear? Third, how do these individuals respond to the discrimination trainings in my paradigm?

3.1 Introduction

Generalization of conditioned fear is exaggerated in anxiety patients (for meta-analyses see Cooper et al., 2022; Fraunfelter, Gerdes, & Alpers, 2022) as well as individuals at risk of anxiety disorders (for a meta-analysis see Sep et al., 2019) compared to healthy controls. Because it seems to be related to an incapacity in perceptually discriminating threat from similar safe stimuli (Holt et al., 2014; Zaman et al., 2019) improving stimuli discrimination is a promising approach to reduce fear generalization. Taking up this idea, some studies developed discrimination trainings, which were able to prevent the occurrence of fear generalization (e.g., Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017). Following the logic of a therapeutic approach, results of Study 1A of this thesis showed that discrimination trainings with fear-relevant stimuli or feedback on perceptual performance reduced generalization especially effective.

However, the sample sizes of the proof-of-principle Study 1A was rather small. One goal of Study 2 (Part A) was therefore to replicate the findings in a bigger sample.

Another limitation of Study 1 was that only healthy participants were examined, who exhibit moderate fear generalization compared to highly anxious individuals or anxiety patients (Cooper et al., 2022; Fraunfelter et al., 2022; Sep et al., 2019). Thus, another goal of Study 2 (Part B) was to consider individual differences in risk factors of anxiety disorders (Lonsdorf & Merz, 2017; Mineka & Zinbarg, 2006) to approximate more clinical samples. Affected individuals have a special need for fear-reducing treatments as they likely show strong generalization of fear (Sep et al., 2019).

3.2 Study 2A: Fear-reducing effect of fear relevance and feedback in discrimination training - a replication test

Study 2A should replicate Study 1A in a larger sample to verify the beneficial role of fear relevance and feedback during discrimination training on the reduction of fear generalization. After all, the fear-reducing effect of fear-relevant training stimuli was rather small in the previous study. Therefore, and also in light of the replication crisis that has been much discussed in psychology for several years (Ioannidis, 2005; LeBel, McCarthy, Earp, Elson, & Vanpaemel, 2018; Pashler & Wagenmakers, 2012), a replication study of the previous findings was advisable.

The replication crisis has arisen because results frequently fail to be replicated, raising doubts about the reliability of published research findings (Pashler & Wagenmakers, 2012). This would mean that a pattern apparently identified in an original study is actually just noise (Maxwell, Lau, & Howard, 2015). Such false-positive findings are favored by low statistical power, for example due to low sample size or small effects (Button et al., 2013). Therefore, when planning a replication study, it is recommended to use a large sample size to ensure sufficient power even for expected small effect sizes (Maxwell et al., 2015). Furthermore, methodological

similarity to the original study should be sufficient (i.e., direct replication) or differences from the original study should at least be considered (LeBel et al., 2018). Ideally, the researchers of the original and replication studies are independent to protect against confirmation bias (LeBel et al., 2018).

The latter suggestion understandably cannot be met by the here presented study, but it allowed repeating the same experiment and analyses with a sample three times as large. Accordingly, every participant underwent the experimental procedure established in Study 1A, consisting of a differential fear conditioning and generalization paradigm as well as a discrimination training (relevant_DT_FB, relevant_DT_noFB, irrelevant_DT_FB, irrelevant_DT_noFB) between the two generalization blocks.

I expected to replicate the training effects of Study 1A, i.e., more effective reduction of fear generalization by the fear-relevant compared to the fear-irrelevant discrimination training, and increased effectiveness of both trainings by reinforcing feedback vs. no feedback.

3.2.1 Materials and methods

3.2.1.1 Participants

The study was conducted within the context of the collaborative research center SFB-TRR-58 project Z02. It was advertised mainly on the Internet, on the website wuewowas.de and a platform of the University of Würzburg. In addition, flyers were posted in public places including university and shops in the vicinity of the research institute. Every person who was interested in participation was screened for exclusion criteria by a telephone interview. Exclusion criteria included left-handedness, non-Caucasian ancestry, use of psychoactive medications, excessive alcohol, nicotine, and caffeine use, use of illegal drugs, severe medical illness, or pregnancy. Absence of

current and/or lifetime DSM-IV diagnoses of axis I mental disorders was assessed using the German version of the Mini International Psychiatric Interview (M.I.N.I., Sheehan et al., 1998). Drug abstinence and pregnancy were assessed using urine screening tests.

In the end, 289 participants were invited to the laboratory, of which 2 were excluded from the analysis due to missing questionnaire data, 3 due to missing rating data and 40 due to missing SCR data or meeting the SCR non-responder criterion (for details see section 3.2.1.5). Accordingly, 244 were included in the statistical analyses. After arrival, all participants read and signed the informed consent and then were randomly assigned to one of four experimental groups (for group descriptives see **Table 5** and **Supplementary Table 7**). There were no pre-experimental differences between groups regarding gender or age. Participants of the feedback groups had lower anxiety scores (for more details on how the score was determined, see section 3.3.1.2) than participants who did not receive feedback ($F(1, 240) = 5.67$, $p = .018$, $\eta_p^2 = .02$; **Table 5**). Analogous differences between feedback vs. no-feedback groups were observed for the Beck-Depression Inventory (BDI II, Hautzinger et al., 2006; $F(1, 240) = 4.94$, $p = .027$, $\eta_p^2 = .02$; **Table 5**). To rule out the possibility that training effects occurred only because of differences in BDI, I repeated all analyses controlling for BDI by including it as covariate. Because the effects remained unchanged, the results of these ANCOVAs are not reported. Group differences in anxiety were controlled for in Study 2B, where anxiety was included as a risk factor.

The study was approved by the ethical committee of the medical board of the University of Würzburg and was conducted in accordance with the ethical principles of the Helsinki Declaration.

3.2.1.2 Screenings and questionnaires

Prior to the experiment, the project Z02 of the collaborative research center SFB-TRR-58 included a clinical interview (Mini-International Neuropsychiatric Interview, M.I.N.I., Sheehan et al., 1998), blood sampling, urine screening tests of pregnancy and drug intake as well as the assessment of several questionnaires. A selection of questionnaires was used to determine the participants' trait anxiety and were considered as risk factor in Study 2B (for details see section 3.3.1.1 and Appendix 6.3.1). All further measures of project Z02 are not part of the research questions of this thesis and therefore are not described in more detail.

Table 5 *Sample characteristics of Study 2*

	relevant _DT _noFB	*relevant _DT _FB*	*irrelevant _DT _noFB*	*irrelevant _DT _FB*	*comparisons*
N	55	64	59	66	
gender (♂)	14	17	18	24	$\chi(3) = 2.18$, $p = .535$
age (*SD*)	23.42 (4.04)	24.11 (5.94)	24.54 (4.79)	24.12 (5.85)	$F(1,240) = 0.68$, $p = .412$
anxiety (*SD*)	0.10 (0.79)	-0.17 (0.73)	0.17 (0.92)	-0.08 (0.92)	$F(1,240) = 5.67$, $p = .018$
BDI (*SD*)	6.00 (5.88)	4.80 (4.40)	7.42 (5.78)	4.98 (6.16)	$F(1,240) = 7.15$, $p = .008$

Notes. DT discrimination training, *FB* feedback, *anxiety* includes the aggregated data of Anxiety Sensitivity Index (ASI-3), Agoraphobic Cognitions Questionnaire and Social Phobia and Anxiety Inventory (SPAI; following the suggestion of Baumann et al., 2017), *BDI* Beck Depression Inventory II, comparisons indicate significantly higher anxiety and BDI in no-feedback compared to feedback groups.

3.2.1.3 Stimulus material

Stimulus material of the conditioning and generalization experiment were identical with Study 1A (see section 2.2.1.3).

3.2.1.4 Procedure and discrimination trainings

The experimental procedure, including the discrimination trainings, was identical with that of Study 1A (see sections 2.2.1.4 and 0). Accordingly, the experiment consisted of a differential fear conditioning and generalization paradigm as well as a discrimination training (*relevant_DT_FB, relevant_DT_noFB, irrelevant_DT_FB, irrelevant_DT_noFB*) between the two generalization blocks.

3.2.1.5 Data recording and reduction

Data recording and reduction happened in the same way as for Study 1A (see section 2.2.1.6). SCR data of some participants (n = 21) could not be used due to a failure of recording technology. The participants concerned were excluded from the sample and in compensation, additional participants were invited to the lab. Moreover, participants with an overall raw mean response smaller 0.02 µS were considered non-responders and therefore excluded from statistical analysis (n = 19).

3.2.1.6 Statistical analysis

Statistical analyses were carried out in the R software environment (version 3.6.1) using packages ‘afex’ (version 0.26-0; Singmann et al., 2015) and ‘emmeans’ (version 1.4.5; Lenth & Lenth, 2018).

The acquisition phase was analyzed with ANOVAs having the within-subject factors stimulus (CS+, CS-) and block, with the factor block having three levels (habituation, acquisition 1, acquisition 2) for arousal, valence and SCR data, and two levels (acquisition 1, acquisition 2) for US-expectancy data.

For analysis of fear generalization pre training, ANOVAs had a within-factor stimulus (CS+, GS1, GS2, GS3, GS4, CS-) only.

Besides, generalization was summarized to one value calculating the Generalization Index (GI = [(GS1 + GS2 + GS3 + GS4)/CS+], for details see Lenaert et al., 2016) separately for generalization pre and post training as described in Study 1 (also see section 2.2.1.7). The GIs were checked for outliers, i.e., values below quartile 1 - 1.5 IQR (inter-quartile range) or above quartile 3 + 1.5 IQR, which would correspond to outliers presented in boxplots. If an individual had outlier GIs for an outcome measure at both time points, and if these were caused by incorrect polarization of the CSs (i.e., CS- > CS+), he or she was excluded from the analysis of the respective outcome measure. Consequently, two participants were removed from analyses of discrimination-training effects for US-expectancy and valence ratings.

Discrimination-training effects were investigated using ANCOVAs on GI post training including fear relevance (relevant_DT, irrelevant_DT) and feedback (with, without) as between-subjects factors and GI pre training as covariate.

The significance level was set at $p < .05$. Greenhouse-Geisser correction of degrees of freedom was applied when the sphericity assumption was violated. Partial eta squares are given for effect size. Simple contrasts corrected for Bonferroni were used for post hoc tests.

3.2.2 Results

3.2.2.1 Acquisition of conditioned fear

Fear acquisition was successful for US-expectancy ratings (**Figure 11** A) as indicated by a significant main effect of stimulus ($F(1, 243) = 794.19$, $p < .001$, $\eta_p^2 = .77$), but not block ($F(1, 243) = 1.48$, $p = .224$, $\eta_p^2 < .01$) or their interaction ($F(1, 243) =$

0.04, $p = .846$, $\eta_p^2 < .01$). ANOVAs on arousal (**Figure 11** B) and valence (**Figure 11** C) ratings returned significant main effects of stimulus (arousal: $F(1, 243) = 323.94$, $p < .001$, $\eta_p^2 = .57$, valence: $F(1, 243) = 93.06$, $p < .001$, $\eta_p^2 = .28$) and block (arousal: $F(1.73, 420.07) = 111.09$, $p < .001$, $\eta_p^2 = .31$, valence: $F(1.73, 419.38) = 33.86$, $p < .001$, $\eta_p^2 = .12$) as well as their interaction (arousal: $F(1.86, 452.76) = 169.85$, $p < .001$, $\eta_p^2 = .41$, valence: $F(1.80, 437.69) = 128.74$, $p < .001$, $\eta_p^2 = .35$). Post-hoc contrasts confirmed that ratings were higher for CS+ vs. CS- regarding arousal (Acquisition1: $F(1, 243) = 261.69$, $p < .001$, $\eta_p^2 = .52$; Acquisition2: $F(1, 243) = 405.76$, $p < .001$, $\eta_p^2 = .63$; Bonferroni corrected $\alpha < .017$), and valence (Acquisition1: $F(1, 243) = 109.33$, $p < .001$, $\eta_p^2 = .31$; Acquisition2: $F(1, 243) = 155.88$, $p < .001$, $\eta_p^2 = .39$; Bonferroni corrected $\alpha < .017$), while no CS+ vs. CS- differences were evident for the habituation block (arousal: $F(1, 243) = 0.6$, $p = .438$, $\eta_p^2 < .01$; valence: $F(1, 243) = 6.45$, $p = .012$, $\eta_p^2 = .03$). These results are indicative of successful fear acquisition.

The Stimulus × Block interaction was not observed for SCR analysis ($F(1.32, 320.03) = 1.92$, $p = .163$, $\eta_p^2 < .01$; **Figure 11** D). Instead, the ANOVA revealed a significant main effect of stimulus ($F(1, 243) = 54.78$, $p < .001$, $\eta_p^2 = .18$) indicating stronger reactions to CS+ vs. CS- during all blocks and a main effect of block ($F(1.43, 347.10) = 69.72$, $p < .001$, $\eta_p^2 = .22$). Post-hoc contrasts of the latter effect showed stronger SCRs during Acquisition1 compared to both habituation ($F(1, 243) = 95.91$, $p < .001$, $\eta_p^2 = .28$) and Acquisition2 ($F(1, 243) = 68.58$, $p < .001$, $\eta_p^2 = .22$), while response strength did not differ between habituation and Acquisition2 ($F(1, 243) = 0.28$, $p = .598$, $\eta_p^2 < .01$).

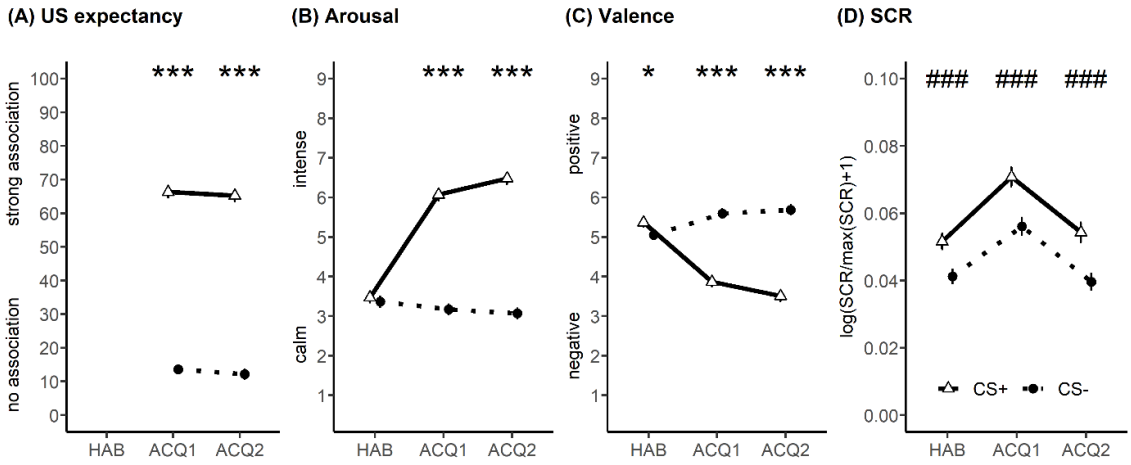


Figure 11. Habituation and acquisition of Study 2.

Notes. Means (with SEs) of US-expectancy (A), arousal (B), and valence (C) ratings and skin conductance responses (SCR; D) are shown for CS+ (white) and CS- (black) for habituation (HAB) and for the acquisition blocks (ACQ1, ACQ2). Significance symbols indicate simple post-hoc contrasts (*** p < .001, * p < .05) or main effects (### p < .001).

3.2.2.2 Generalization of conditioned fear

Significant main effects of stimulus were revealed for US-expectancy (F(2.95, 717.52) = 327.86, p < .001, η_p^2 = .57; **Figure 12** A), arousal (F(3.25, 789.86) = 244.42, p < .001, η_p^2 = .50; **Figure 12** B), and valence (F(3.23, 784.31) = 148.37, p < .001, η_p^2 = .38; **Figure 12** C) ratings as well as SCR (F(3.42, 832.05) = 42.37, p < .001, η_p^2 = .15; **Figure 12** D). Simple contrasts (Bonferroni corrected, α < .01) indicate higher US expectancy (F(1, 243) = 755.17, p < .001, η_p^2 = .76), higher arousal (F(1, 243) = 579.12, p < .001, η_p^2 = .70), more negative valence (F(1, 243) = 277.38, p < .001, η_p^2 = .53) and stronger SCR (F(1, 243) = 84.54, p < .001, η_p^2 = .26) for CS+ than for CS-. Compared to CS-, fear generalization, i.e., increased responses, was observed for all outcome measures to GS1 (US expectancy: F(1, 243) = 337.7, p < .001, η_p^2 = .58; arousal: F(1, 243) = 319.12, p < .001, η_p^2 = .57; valence: F(1, 243) = 219.41, p < .001, η_p^2 = .47; SCR: F(1, 243) = 52.02, p < .001, η_p^2 = .18), and GS2 (US expectancy: F(1, 243) = 88.4, p < .001, η_p^2 = .27; arousal: F(1, 243) = 154.3, p < .001, η_p^2 = .39; valence: F(1, 243) =

86.23, $p < .001$, $\eta_p^2 = .26$; SCR: $F(1, 243) = 14.93$, $p < .001$, $\eta_p^2 = .06$). On rating level only, fear was also generalized to GS3 (US expectancy: $F(1, 243) = 42.35$, $p < .001$, $\eta_p^2 = .15$; arousal: $F(1, 243) = 95.75$, $p < .001$, $\eta_p^2 = .28$; valence: $F(1, 243) = 51.91$, $p < .001$, $\eta_p^2 = .18$; SCR: $F(1, 243) = 6.47$, $p = .012$, $\eta_p^2 = .03$) and GS4 (US expectancy: $F(1, 243) = 12.05$, $p < .001$, $\eta_p^2 = .05$; arousal: $F(1, 243) = 27.93$, $p < .001$, $\eta_p^2 = .10$; valence: $F(1, 243) = 21.24$, $p < .001$, $\eta_p^2 = .08$; SCR: $F(1, 243) = 0.17$, $p = .677$, $\eta_p^2 < .01$).

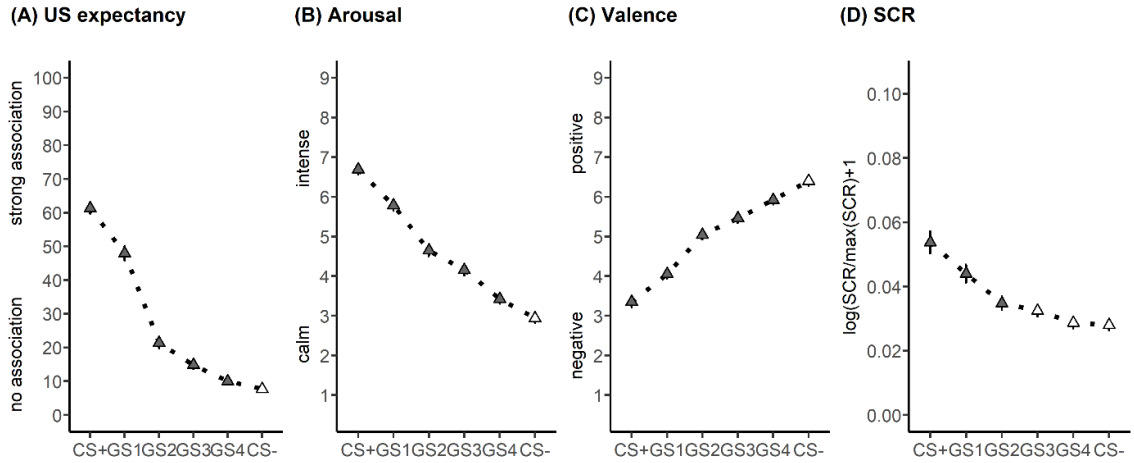


Figure 12. Generalization gradients pre training.

Notes. Means (with SEs) of ratings of US expectancy (A), arousal (B), and valence (C), and skin conductance responses (SCR; D) are shown for CS+, CS- as well as GS1-GS4. Gray shaded data points indicate a significant difference compared to CS-.

3.2.2.3 Discrimination training effects

3.2.2.3.1 Planned analyses

For US-expectancy ratings (the corresponding generalization gradients are depicted in **Figure 13** A-D), the ANCOVA on GIs revealed a significant main effect of feedback ($F(1, 237) = 3.9$, $p = .049$, $\eta_p^2 = .02$; **Figure 14** A), but no effect of fear relevance ($F(1, 237) = 0.3$, $p = .585$, $\eta_p^2 < .01$) or Fear relevance × Feedback interaction ($F(1, 237) = 0.06$, $p = .802$, $\eta_p^2 < .01$; **Figure 13** E), indicating that fear generalization

was reduced more effectively by rewarding feedback conditions, while fear relevance of the training stimuli did not increase effectiveness.

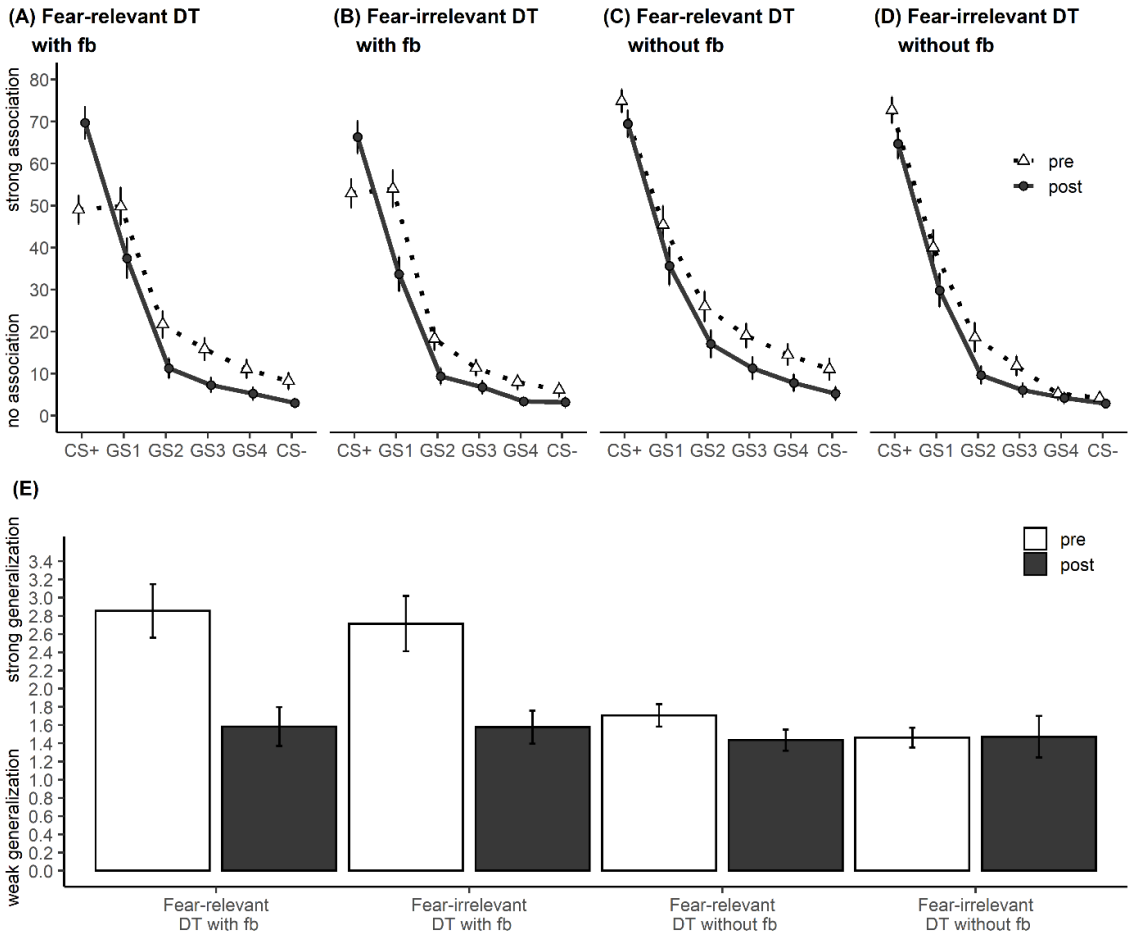


Figure 13. Generalization of US expectancy divided by training group.

Notes. DT discrimination training, *fb* feedback, depicted are Means with SEs of gradients (A - D) and Generalization Indices (GI; E) pre and post training, respectively.

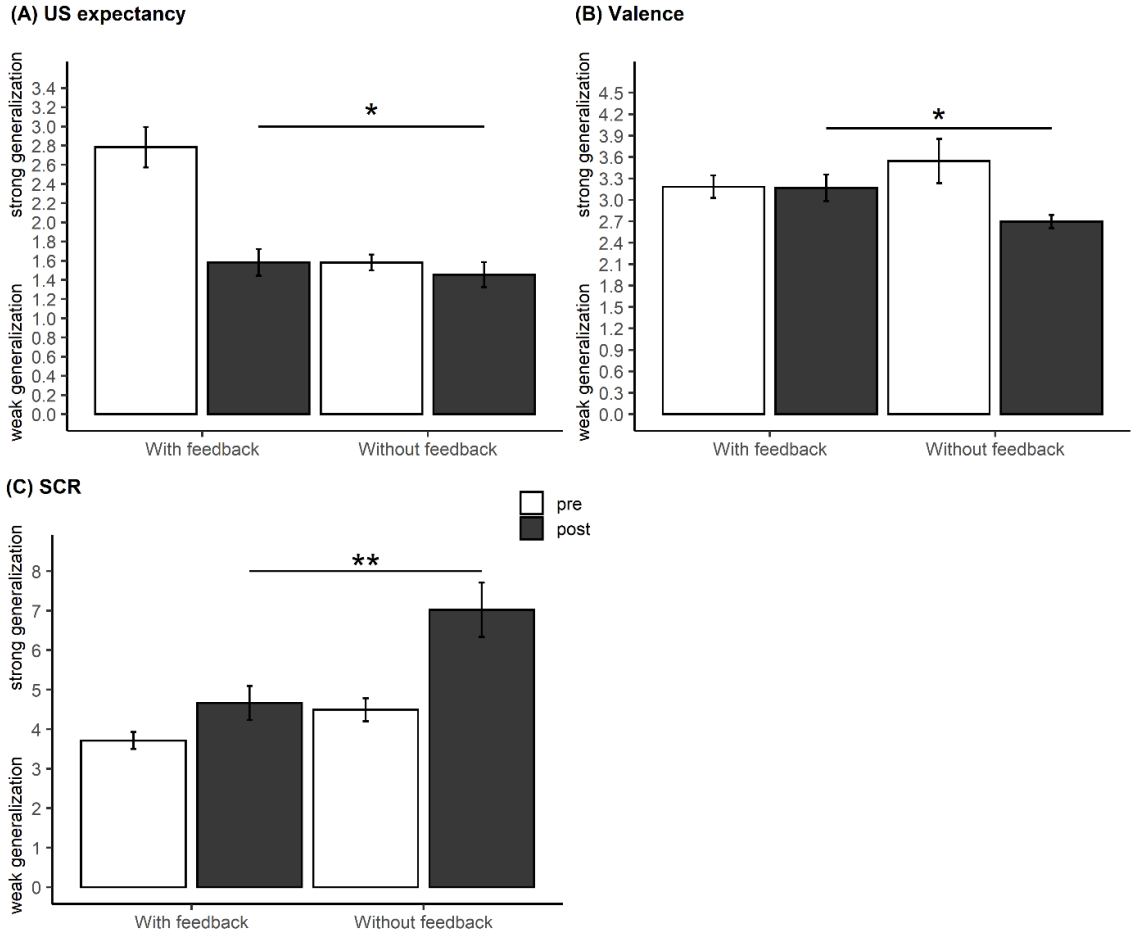


Figure 14. Generalization indices (GI) averaged by groups with vs. without feedback.

Notes. Bar graphs depict means with SEs for US expectancy (A) valence (B) and SCR (C) for generalization pre and post training, significance symbols indicate main effects of ANCOVA, * p < .05., ** p < .01.

Analysis of arousal ratings (the corresponding generalization gradients are depicted in **Figure 15** A-D) with ANCOVA on GIs showed no effects involving the between-factors (all p values > .306, **Figure 15** E).

For valence ratings (the corresponding generalization gradients are depicted in **Figure 16** A-D), the ANCOVA on GIs had a significant main effect of feedback (F(1, 237) = 5.95, p = .015, η_p^2 = .02; **Figure 14** B), but no further between-factor effects (fear relevance: F(1, 237) = 0.01, p = .932, η_p^2 < .01; Fear relevance × Feedback: F(1,

237) = 0.37, p = .543, η_p^2 < .01; **Figure 16** E). According to the result, feedback vs. no-feedback groups showed stronger fear generalization post training.

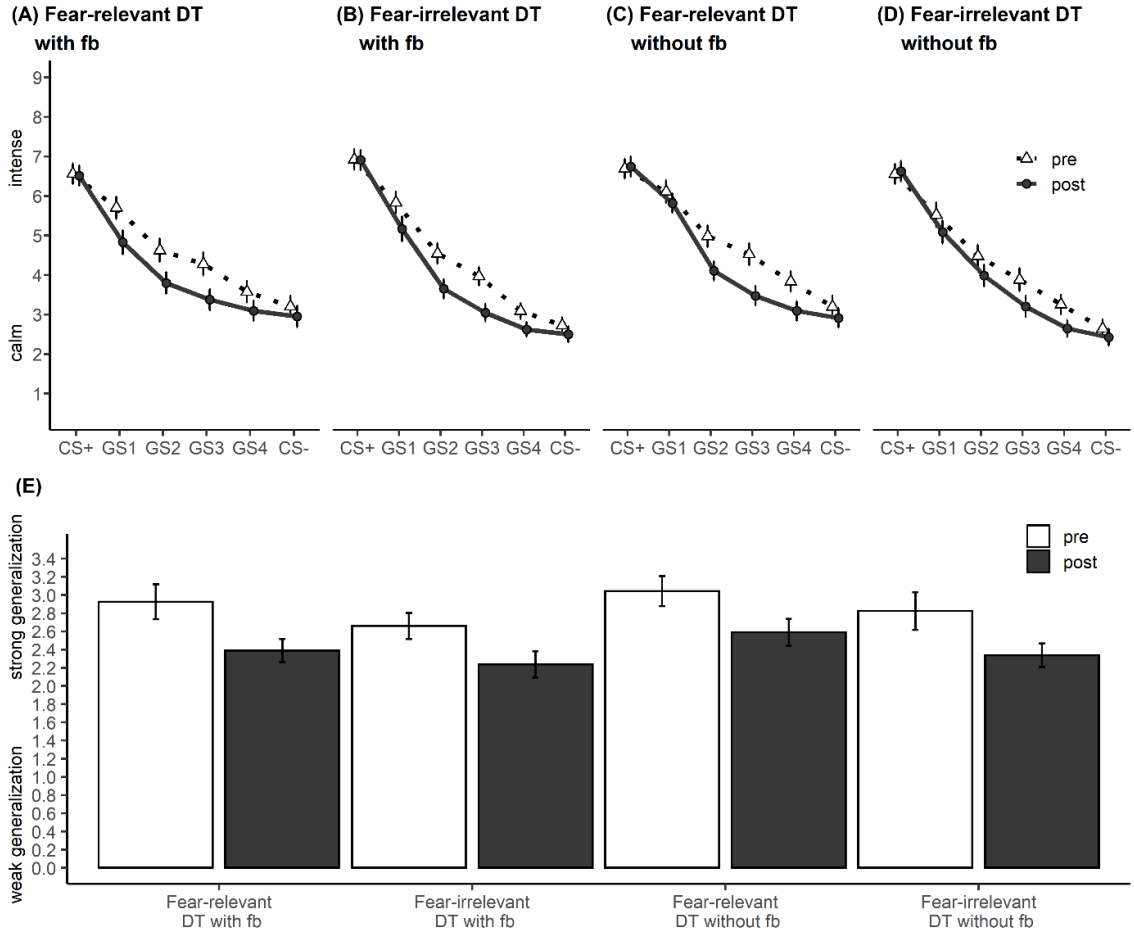


Figure 15. Generalization of arousal divided by training group.

Notes. DT discrimination training*, fb* feedback, depicted are Means with SEs of gradients (A - D) and Generalization Indices (GI; E) pre and post training, respectively.

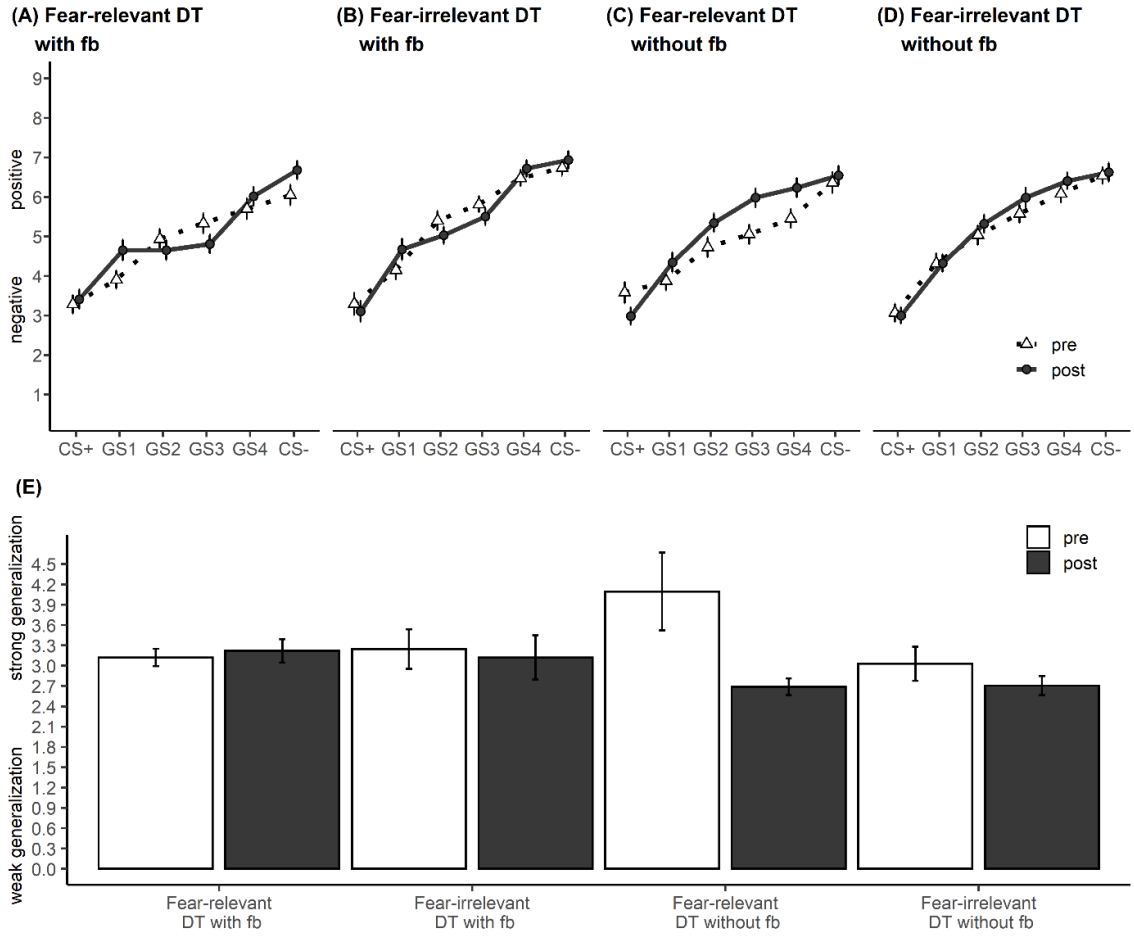


Figure 16. Generalization of valence divided by training group.

Notes. DT discrimination training*, fb* feedback, depicted are Means with SEs of gradients (A - D) and Generalization Indices (GI; E) pre and post training, respectively.

Analysis of SCR with ANCOVA on GIs (the corresponding generalization gradients are depicted in **Figure 17** A-D) revealed a significant main effect of feedback ($F(1, 239) = 6.87$, $p = .009$, $\eta_p^2 = .03$; **Figure 14** C), but no further between-factor effects (fear relevance: $F(1, 239) = 0.01$, $p = .928$, $\eta_p^2 < .01$; Fear relevance × Feedback: $F(1, 239) = 0.51$, $p = .478$, $\eta_p^2 < .01$; **Figure 17** E), indicating that fear generalization was reduced more effectively by rewarding feedback conditions, while fear relevance of the training stimuli did not increase effectivity.

The main effect of covariate was significant for all outcome measures (US expectancy: $F(1, 237) = 74.56$, $p < .001$, $\eta_p^2 = .24$; arousal: $F(1, 239) = 36.06$, $p < .001$,

η_p^2 = .13; valence: $F(1, 237) = 8.83$, $p = .003$, η_p^2 = .04, SCR: $F(1, 239) = 6.42$, $p = .012$, η_p^2 = .03) suggesting positive associations between generalization assessed pre and post training (US expectancy: $r(240) = .48$, $p < .001$; arousal: $r(242) = .37$, $p < .001$; valence: $r(240) = .17$, $p = .006$, SCR: $r(242) = .18$, $p = .004$).

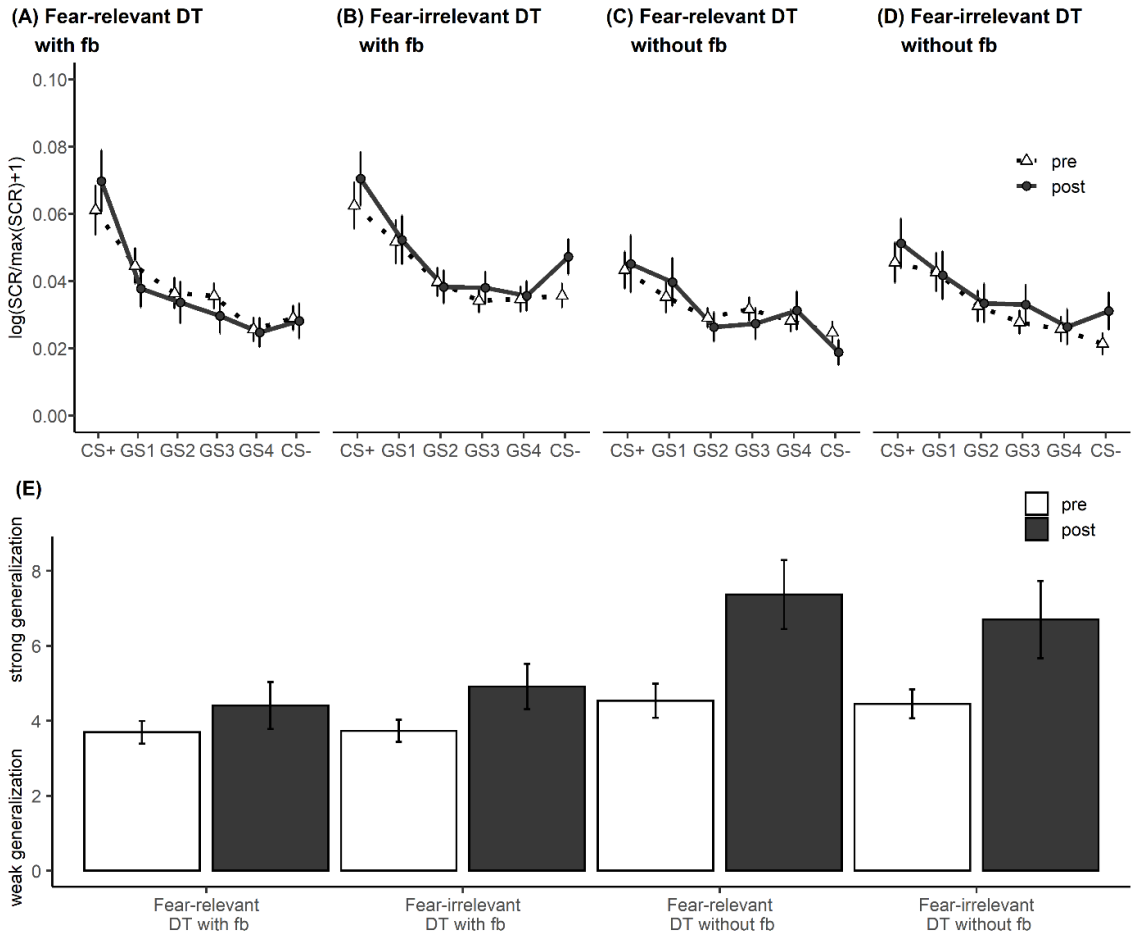


Figure 17. Generalization of skin conductance response (SCR) divided by training group.

Notes. DT discrimination training, *fb* feedback, depicted are Means with SEs of gradients (A - D) and Generalization Indices (GI; E) pre and post training, respectively.

3.2.2.3.2 Additional post-hoc analyses

For a better understanding of the discrimination-training effects, I performed some post-hoc analyses.

First, I checked whether the two feedback groups differed with regard to generalization (i.e., GI) already pre training. For this purpose, I performed post-hoc

ANOVAs on GI pre of US-expectancy ratings, valence ratings and SCR having the between-subjects factor feedback (with, without).

The ANOVA on GI pre of US-expectancy rating returned a significant main effect of feedback ($F(1, 240) = 25.97$, $p < .001$, $\eta_p^2 = .10$), indicating that GI pre training was higher in the feedback than the no-feedback condition. The same ANOVA on valence ratings revealed equal levels of GI pre training ($F(1, 240) = 1.14$, $p = .288$, $\eta_p^2 < .01$). In contrast, for SCR, participants assigned to the feedback condition had lower GIs pre training compared to participants in the no-feedback condition, as indicated by a significant main effect of feedback ($F(1, 242) = 4.78$, $p = .030$, $\eta_p^2 = .02$).

Second, I examined for US-expectancy ratings if the feedback and no-feedback group already differed during learning phase, running a post-hoc ANOVA on the CS difference of US-expectancy ratings at acquisition 2 with a between-subjects factor feedback (with, without).

The ANOVA on CS difference of US-expectancy ratings at acquisition 2 returned a significant main effect of feedback ($F(1, 240) = 102.30$, $p < .001$, $\eta_p^2 = .30$), indicating that the feedback and no-feedback group already differed at the end of acquisition phase therein that the feedback group's CS difference was smaller.

Lastly, I was interested in how contingency awareness (as reflected in CS difference of US-expectancy ratings) post training is related with the extent of generalization (reflected in GI) post training across outcome measures (i.e., US-expectancy and valence ratings and SCR), and therefore computed Pearson's correlations.

Significant correlations suggested negative associations between contingency awareness (as reflected in CS difference of US-expectancy ratings) post training and

generalization (i.e., GI) post training (US expectancy: $r(240) = -.56$, $p < .001$; valence: $r(240) = -.13$, $p = .039$, SCR: $r(242) = -.15$, $p = .023$).

3.2.3 Discussion

Study 1A, was the first to come up with a design that allowed to verify discrimination-training effects by comparing generalization indices pre vs. post training. In the first place, Study 2 was conducted to replicate the findings of this proof-of-principle study. In accordance with recommendations for replication studies (LeBel et al., 2018; Maxwell et al., 2015), I re-used the same paradigm and analyses in a large sample and tested the hypothesis that the generalization of conditioned fear to perceptually similar stimuli is significantly better reduced by a discrimination training with fear-relevant vs. fear-irrelevant stimuli. A second hypothesis concerned feedback, where I expected that discrimination-training effects benefit by reinforcing feedback.

To examine the role of the discrimination training on fear generalization, I first examined whether conditioned fear was acquired and how this conditioned fear generalized to other stimuli (i.e., the generalization gradient). Analysis of acquisition data is indicative of successful fear conditioning on rating level as the threat cue (CS+) vs. the safety cue (CS-) was rated more arousing, more negative, and was stronger associated with the US. Moreover, the CS+ triggered larger SCRs as compared to the CS-. However, a differential SCR response could already be observed during habituation trials, which is why it cannot for sure be interpreted as result of fear learning. Further, these conditioned fear responses remained stable during generalization test pre discrimination training. Moreover, the generalization test revealed generalization of conditioned fear to some of the morph stimuli. While larger SCRs were triggered by the two GSs most similar to CS+ (i.e., GS1 and GS2), the ratings are indicative for increased fear to all GSs, compared to CS- respectively.

Overall, these results are consistent with previous studies using similar paradigms (Holt et al., 2014; Lissek et al., 2008; Schiele et al., 2016) and indicate successful acquisition of fear to the CS+ as well as its generalization to similar stimuli. Also in line with these previous studies, fear was generalized to more stimuli at the rating level than at the physiological level (i.e., SCR).

Analyses of discrimination-training effects on fear generalization confirmed stronger effects for discrimination training with feedback vs. without feedback. However, in contrast to the previous study, no support was found for stronger effects for the fear-relevant vs. the fear-irrelevant discrimination training.

There were valid reasons for my hypothesis that a reduction of fear generalization is especially effective when the discrimination training is carried out with fear-relevant stimuli vs. fear-irrelevant stimuli. In addition to the proof-of-principle study presented here (Study 1A; K. Herzog et al., 2021), these include, for example, previous reports that preventive discriminative training with fear-relevant stimuli can reduce avoidance behavior toward generalization stimuli (Lommen et al., 2017). Moreover, there is more general evidence on the transfer of discrimination learning to a test phase, which was found to be more effective when relevant stimuli were trained compared to irrelevant ones (Furmanski & Engel, 2000). Nevertheless, in contrast to the previous study, I must reject the hypothesis based on the present results. Fear relevance effects were already low in the proof-of-principle study (Study 1A). One explanation is that discrimination training, also with fear-irrelevant stimuli, can reduce subsequent generalization effects in general (Ginat-Frolich et al., 2019; Ginat-Frolich et al., 2017). To some extent, this seems to contradict the findings that perceptual learning is very specific and cannot necessarily be transferred to other stimulus material (Dosher & Lu, 2017; Furmanski & Engel, 2000). However, several mechanisms are involved in

perceptual learning, including sensory processing, decision-making and learning processes, attention, and feedback (Dosher & Lu, 2017). They are all triggered during discrimination training, regardless of whether it is fear-relevant or fear-irrelevant, and could contribute to fear reduction in the subsequent generalization test. The involvement of higher-order processes in fear generalization and its reduction could also explain why generalized fear is more pronounced in higher-processed measures such as ratings than in SCR.

The present study supports the conclusion that reinforcing feedback compared to no feedback during discrimination training helps to reduce fear generalization. While in the proof-of-principle study (Study 1A) this conclusion was suggested only by the US-expectancy ratings, in the present study it is also supported by physiological data (i.e., SCRs). For both outcome measures, fear generalization after training is reduced in the groups with feedback compared to the groups without feedback. Obviously, the two effects are related. Previous studies have shown that SCR depends on contingency awareness (Constantinou et al., 2021; Hamm & Vaitl, 1996; Lipp & Purkis, 2005; but see Schultz & Helmstetter, 2010). This also seems to be the case for the present data, as suggested by the exploratory post-hoc correlation of CS differentiation (for US expectancy) and generalization index (for SCR), each after training. However, I do not want to hide the fact that the effect for SCR is not as clear as for the US-expectancy ratings, since GI of SCR increased pre to post training (as already reported in Study 1). Contradictory results were seen for valence ratings, where fear generalization was lower after training in the conditions without feedback than in the training conditions with feedback. It is conceivable that this effect is attributable to the wide range and large variance of the GIs for valence ratings (see **Figure 19** G of Study 2B).

The results of Study 1B did not support the assumption that the implemented reinforcing feedback improved mood (Westermann, Spies, Stahl, & Hesse, 1996), attention (Chelazzi, Perlato, Santandrea, & Della Libera, 2013) and motivation (Fishbach, Eyal, & Finkelstein, 2010) of the participants in general, which consequently improved discrimination during the subsequent generalization test. Otherwise, feedback should have had a fear-reducing effect whether it was administered in discrimination training or non-discriminative control training (see Appendix 6.2.2.2). Instead, it is more likely that feedback has a specific effect on discrimination learning during training. However, when analyzing discrimination performance during training in the present study (see Appendix 6.3.2), no direct support for this assumption is initially found. Specifically, there is no interaction of feedback and discrimination improvement from part 1 to part 2 of the training. Instead, feedback leads to better discrimination performance in both parts of the training. The reason could be that feedback has an effect on discrimination performance right from the beginning of the training, as also reported in previous studies on perceptual learning (Dosher & Lu, 2017; M. H. Herzog & Fahle, 1999). Based on these findings, the specific effect of feedback on discrimination learning simply occurred very quickly.

However, I do not want to hide that the effect might have been favored by group differences pre training. Additional post-hoc tests for US-expectancy ratings revealed that the feedback groups showed greater generalization (i.e., higher GI scores) before training and poorer CS differentiation at the end of fear acquisition compared to the no-feedback groups. The greater need for training makes its success more likely. The effect is favored by the principle of initial values, which states that high values have a greater potential to decrease, whereas low values decrease only slightly as they approach the minimum (Jennings & Cribbie, 2016).

Therefore, on the one hand, these group differences pre training are a limitation of Study 2. On the other hand, they provide an opportunity to distinguish between individuals with different treatment needs and are therefore a strength of the sample when accounted for. I did this in Study 2B to examine whether risk groups with a high need for fear-reducing treatment (e.g., due to a deficit in threat-safety learning) can particularly benefit from discrimination training. This will also allow me to obtain preliminary indications of the effect of training discrimination with more clinically relevant individuals who are known to recognize safety signals more poorly (Lissek et al., 2009) and generalize fear more strongly (Laufer, Israeli, & Paz, 2016; Lissek et al., 2014; Lissek et al., 2010; Sep et al.,2019).

3.3 Study 2B: Fear generalization and responsiveness to discrimination training depending on individual risk for anxiety disorder

Study 2B aimed to extend the findings of Studies 1A and 2A by examining individual characteristics and their influence on fear generalization and its reduction. Of particular interest were risk factors of anxiety disorders (Lonsdorf & Merz, 2017; Mineka & Zinbarg, 2006) that come along with strong generalization of fear (Sep et al., 2019), because affected individuals have a special need for fear-reducing interventions and could particularly benefit from a discrimination training.

A much studied vulnerability factor among personality traits is subclinical trait anxiety (Raymond et al., 2017). It has already been shown to promote overgeneralization of conditioned fear (Baumann et al., 2017; Stegmann et al., 2019; but see Torrents-Rodas et al., 2013; for a meta-analysis see Sep, Steenmeijer, & Kennis, 2019). Specifically, the study by Baumann and colleagues (2017) used a principal component analysis of several questionnaires to describe the construct anxiety. The

identified factor was characterized by the Anxiety Sensitivity Index (ASI 3; Taylor et al., 2007), the Agoraphobic Cognitions Questionnaire (ACQ; Chambless et al., 1984) and the Social Phobia and Anxiety Inventory (SPAI; Turner et al., 1989). Two extreme groups of anxiety-factor values were compared in terms of their fear response to an aversive conditioned face picture (CS+) and four morph stimuli (GS), ranging in perceptually similarity from CS+ to CS- in 20 % steps. The high anxious group showed broader fear generalization for valence and US-expectancy ratings compared to the low anxious group (for a more detailed study description see section 1.3.3).

The present study aimed at confirming the relation between higher trait anxiety and strong fear generalization in a big sample. Different than in the study by Baumann and colleagues (2017), here the relationship between anxiety and generalization was determined on an individual level. This approach emphasizes the view of anxiety as an individual trait, which likely is related to individual fear generalization and responsiveness to fear-reducing treatments, and expands the idea to conceptualize psychopathology as a continuum (Cuthbert, 2014; Insel et al., 2010) to a subclinical level.

Another risk factor of anxiety disorders are individual differences in fear learning (Beckers et al., 2013; Mineka & Oehlberg, 2008; Mineka & Zinbarg, 2006). Anxiety patients show abnormalities in fear vs. safety learning characterized by less differentiation of threat (CS+) and safety signals (CS-, Duits et al., 2015; Lissek et al., 2005). However, to my knowledge, the impact of deficient differentiation of CSs at the end of fear conditioning on the extent of fear generalization to similar neutral cues has not been studied yet. A paired occurrence of weak CS differentiation with strong fear generalization was observed in patient (Lissek, Kaczkurkin, et al., 2014; Lissek et al., 2010) and vulnerable groups (e.g. children; Schiele et al., 2016), but little is known

about the direct connection of the two mechanisms. The latter study reported steeper generalization gradients (for arousal and valence ratings) in healthy adults and children who were contingency aware (as compared to unaware) during generalization (Schiele et al., 2016). Another study showed that certain characteristics of a generalization pattern (i.e., range between CSs) have been already established during fear learning, supporting the idea that differences in generalization may be related to basic characteristics such as the efficacy of conditioning (Stegmann et al., 2019). The present study wants to follow up on these findings and examine the connection between fear learning and fear generalization on an individual level.

For this purpose, this study considered individual differences in threat vs. safety learning at the end of fear conditioning distinguishing between learning on cognitive and affective level. These levels of learning are reflected in US-expectancy and arousal ratings, respectively (Kleinginna & Kleinginna, 1981; Lonsdorf et al., 2017). According to dual-process models these are distinct levels of fear learning (LeDoux & Pine, 2016; Ohman & Mineka, 2001) and therefore might reflect different aspects of deficient threat-safety differentiation.

As a next step, this study addressed the question of whether individual differences in anxious personality and threat vs. safety learning alter the response to fear-reducing discrimination training and which training condition is particularly effective for high-risk participants. To date, fear-reducing discrimination trainings have only been studied for effectiveness at the group level. For example, a group of spider phobics who completed discrimination training developed by Ginat-Frolich and colleagues (2019; 2017) showed lower fear generalization after training than a group of phobics who did not undergo discrimination training.

The inclusion of the individual level, as planned in the present study, provides additional information, using spider phobics as an example, on whether training is more effective for individuals with severe or moderate spider phobia. Such knowledge could allow us to predict the success of different treatments, paving the way for more individualized therapy (S. G. Hofmann & Hayes, 2019).

To examine these questions, participants' anxiety was assessed prior to the experiment using the ASI 3, ACQ, and SPAI questionnaires.

I hypothesized that fear generalization would increase with increasing individual vulnerability, i.e., higher anxiety and lower CS differentiation. Moreover, I expected good responsiveness to fear-relevant discrimination training regardless of individuals' vulnerability. In contrast, I expected that the more vulnerable an individual was (i.e., the more anxious and the less able to differentiate CSs), the less responsive they would be to the fear-irrelevant discrimination training. Similarly, I expected all participants to respond comparably to training with feedback, whereas responsiveness to training without feedback would decrease with increasing vulnerability. No a priori hypotheses were made about differential generalization risk or training response between these two levels.

3.3.1 Materials and methods

The materials and methods are identical to those used in Study 2A. Three of the assessed questionnaires were used to determine participants' trait anxiety and were included as risk factor in this part of Study 2. The selection of the questionnaires and the processing of their scores are described in the following sections.

3.3.1.1 Questionnaires

Within the project Z02 of the collaborative research center SFB-TRR-58 participants completes a series of questionnaires, including a questionnaire on demographic information, e.g., age and gender, and the German versions of several standardized questionnaires. Same as in Study 1, they included the BDI, Trait and state versions of the STAI as well as PANAS (for a description of the questionnaires see section 2.2.1.2). In the current study, I chose three questionnaires to determine the trait anxiety of individuals, all of which assess general and specific anxiety symptoms, i.e., the Anxiety Sensitivity Index (ASI 3; Taylor et al., 2007), the Agoraphobic Cognitions Questionnaire (ACQ; Chambless et al., 1984) and the Social Phobia and Anxiety Inventory (SPAI; Turner et al., 1989).

The German version of the ASI-3 (Kemper, Ziegler, & Taylor, 2009) comprises 18 items assessing fear of anxiety-related sensations. The items are rated on a 5-point Likert scale ranging from zero (“very few”) to four (“very much”) and then combined into a sum score.

The 14 items of the ACQ (Chambless et al., 1984) capture the frequency of thoughts about negative consequences of experiencing anxiety. The items are rated on a five-point Likert scale from 1 (“the thought never occurs”) to 5 (“the thought always occurs”).

The SPAI (Turner et al., 1989) covers cognitive, somatic, and behavioral dimensions of social anxiety. The German version (Fydrich, 1999) has 22 items rated on a seven-point Likert scale from 0 ("never") to 6 ("always").

Previous research has shown that these three questionnaires describe the construct anxiety (Baumann et al., 2017), which is why they were also combined into one score in this study and used as risk factor anxiety (for details see section 3.3.1.2).

3.3.1.2 Statistical analysis

Statistical analyses were performed in the R software environment (version 3.6.1) using packages 'afex' (version 0.26-0; Singmann et al., 2015), 'emmeans' (version 1.4.5; Lenth & Lenth, 2018) and 'lsr' (version 0.5; Navarro, 2015).

The definition of the risk factor trait anxiety was inspired by the study of Baumann et al. (2017), who identified a factor describing the construct anxiety, characterized by the ASI (Taylor et al., 2007), the ACQ (Chambless et al., 1984) and the SPAI (Turner et al., 1989). Following this finding, in this study, each participant's anxiety was defined as the mean of the three questionnaire total scores, which were previously *z*-standardized. As risk factor threat vs. safety differentiation at the cognitive or affective level (i.e., cCSdiff or aCSdiff), the CS difference at the end of acquisition Block 2 of the US-expectancy or arousal ratings was used.

To investigate the influence of the risk factors on fear generalization pre training, ANCOVAs were performed separately for each risk factor, with a within-subject factor stimulus (CS+, GS1, GS2, GS3, GS4, CS-) and the respective risk factor as covariate.

Besides, generalization was summarized into one value by calculating the Generalization Index (GI = [(GS1 + GS2 + GS3 + GS4)/CS+], for details see Lenaert et al., 2016) as described before (see section 2.2.1.7). As in Study 2A, the participants who met my outlier criterion were removed from analyses of the effects of discrimination training (n = 2 for US-expectancy, and n = 2 for valence ratings).

The influence of the risk factors on the effects of discrimination training was examined with linear regression models. They provide the opportunity to examine an interaction between categorical (i.e., training conditions) and continuous (i.e., risk factors) between-subjects factors. To this end, ANCOVAs on GI post training from

Study 1A were first transferred into linear regression models. These models (model A) were set up with the function lm(), with the factors GI pre training, fear relevance (relevant_DT, irrelevant_DT), feedback (with, without), and the interaction Fear relevance × Feedback. Then, models B, C, and D were set up, which included all effects of model A and were expanded by a risk factor (i.e., anxiety, cCSdiff, or aCSdiff) and its interaction with fear relevance and/or feedback.

Model A:

GI post training = GI pre training + fear relevance*feedback

Model B:

GI post training = GI pre training + fear relevance*feedback*anxiety

Model C:

GI post training = GI pre training + fear relevance*feedback*cCSdiff

Model D:

GI post training = GI pre training + fear relevance*feedback*aCSdiff

Pairwise comparison of Model A with one other (Model B, Model C or Model D) using the anova() statement allowed me to see if the inclusion of a risk factor explained variance beyond pre-training GI and training conditions. In case of significant model comparison, I concluded that the respective risk factor added explanatory value and performed post-hoc regression analysis for significant interaction effects involving risk factor split by the involved between-subjects factor.

The significance level was set at $p < .05$. For ANCOVAs, Greenhouse-Geisser correction of degrees of freedom was applied when the sphericity assumption was

violated, simple contrasts (Bonferroni corrected) and Pearson correlations were used as post-hoc tests, and partial eta squares are reported for effect size.

3.3.2 Results

3.3.2.1 Impact of anxious personality on fear generalization

The addition of anxiety as covariate to analyses of fear generalization revealed a main effect of anxiety for all outcome measures except SCR (US expectancy: $F(1, 242) = 4.73$, $p = .031$, $\eta_p^2 = .02$; arousal: $F(1, 242) = 15.62$, $p < .001$, $\eta_p^2 = .06$; valence: $F(1, 242) = 8.68$, $p = .004$, $\eta_p^2 = .03$; SCR: $F(1, 242) = 0.05$, $p = .826$, $\eta_p^2 < .01$; for extreme group generalization gradients see **Figure 18** A-D), whereas the Anxiety × Stimulus interaction did not reach significance for any of the outcome measures (all p-values > .479). Accordingly, higher anxiety was related to higher US expectancy ($r(242) = .14$, $p = .031$, **Figure 18** E), more arousal ($r(242) = .25$, $p < .001$, **Figure 18**F) and more negative valence ($r(242) = .19$, $p = .004$, **Figure 18** G) to all stimuli.

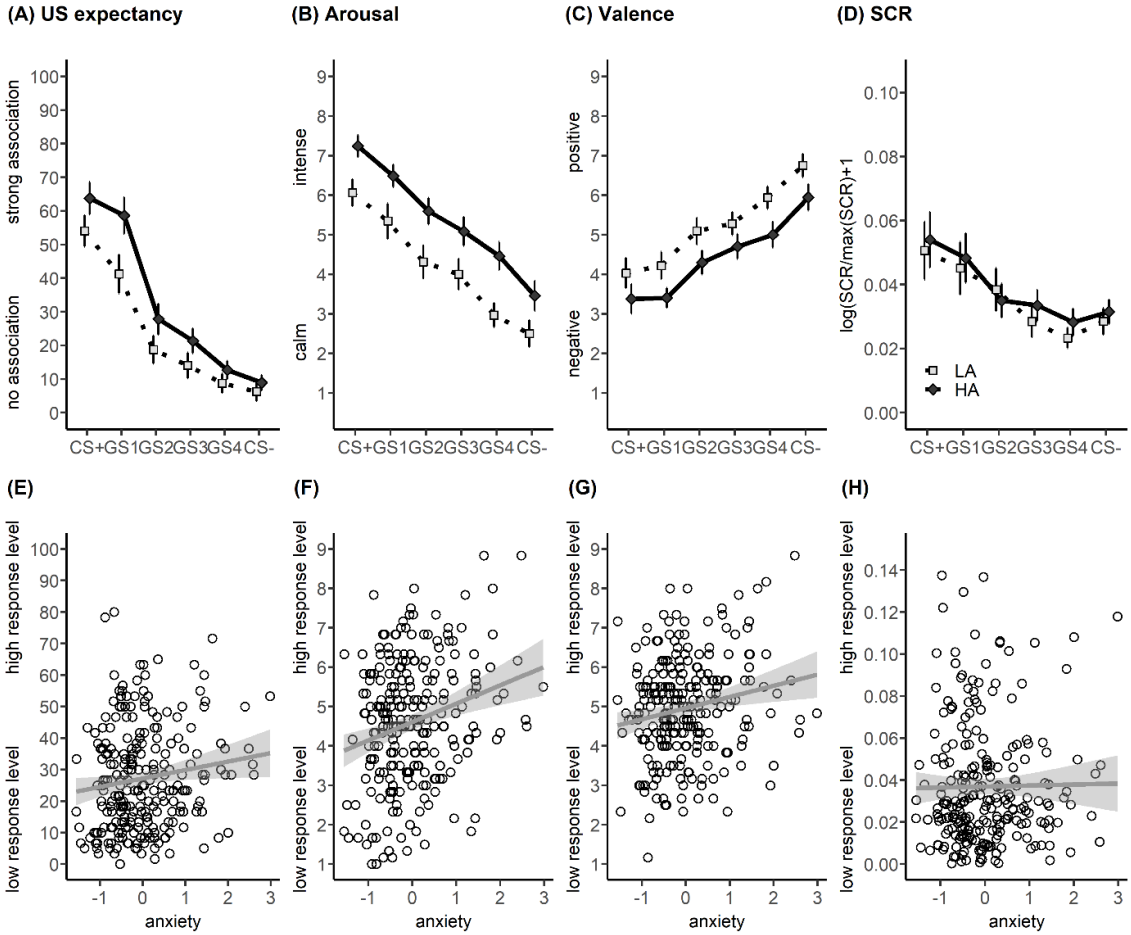


Figure 18. Pre-training generalization gradients and mean response level as a function of anxiety.

Notes. The gradients show means (with SEs) for ratings of US expectancy (A), arousal (B), and valence (C), and skin conductance responses (SCR; D) of extreme groups with low (LA) and high (HA) anxiety, i.e., with anxiety scores that are more than one SD below or above the mean, respectively. Anxiety was associated with higher mean response level for all ratings (E,F,G), but not SCR (H).

3.3.2.2 Impact of impaired fear vs. safety learning on fear generalization

Better cognitive differentiation of threat and safety CS (cCSdiff) at the end of conditioning phase was related to a lower fear response level to all stimuli during generalization phase as shown by a significant main effect of the covariate cCSdiff for US-expectancy ($F(1, 242) = 13.73$, $p < .001$, $\eta_p^2 = .05$) and arousal ($F(1, 242) = 5.01$, $p = .026$, $\eta_p^2 = .02$) ratings, but not for the other dependent variables (all other p-values > .101). Moreover, a significant cCSdiff × Stimulus interaction was revealed for these two

outcome measures (US expectancy: $F(2.93, 709.96) = 29.91$, $p < .001$, $\eta_p^2 = .11$; arousal: $F(3.26, 789.12) = 3.29$, $p = .017$, $\eta_p^2 = .01$; all other p-values > .055; for extreme group generalization gradients see **Figure 19** A-D). Post-hoc correlations of cCSdiff and generalization index GI (**Figure 19** E-H) confirmed that weaker CS differentiation on cognitive level is associated with stronger fear generalization to the GSs, however for US expectancy only ($r(242) = -.42$, $p < .001$), but not arousal ($r(242) = -.04$, $p = .516$). Since the GI disregards the CS-, additional correlations of the risk factor with the CS difference during generalization were calculated, showing that the interaction effect can be explained by a positive influence of the risk factor on the differentiation of the CSs (US expectancy: $r(242) = .53$, $p < .001$; arousal: $r(242) = .18$, $p = .006$).

Better affective differentiation of threat and safety CS (aCSdiff) at the end of conditioning phase, was related to a lower fear response level to all stimuli as indicated by a significant main effect of the covariate aCSdiff for US expectancy only (US expectancy: $F(1, 242) = 4.47$, $p = .035$, $\eta_p^2 = .02$; all other p-values > .257). Moreover, a significant aCSdiff × Stimulus interaction was revealed for all outcome measures (US expectancy: $F(2.97, 717.69) = 3.92$, $p = .009$, $\eta_p^2 = .02$; arousal: $F(3.78, 915.52) = 57.59$, $p < .001$, $\eta_p^2 = .19$; valence: $F(3.62, 875.85) = 34.11$, $p < .001$, $\eta_p^2 = .12$; for extreme group generalization gradients see **Figure 20** A-C) except SCR ($F(3.44, 831.64) = 1.25$, $p = .289$, $\eta_p^2 < .01$; for extreme group generalization gradients see **Figure 20** D). Post-hoc correlations of aCSdiff and generalization index GI (**Figure 20** E-H) confirmed that weaker CS differentiation on affective level is associated with stronger fear generalization to the GSs, however for arousal only ($r(242) = -.30$, $p < .001$), but not US-expectancy ($r(242) = -.07$, $p = .290$) or valence ratings ($r(242) = -.08$, $p = .229$). Additional correlations of the risk factor with the CS difference during

generalization revealed a positive association (US expectancy: $r(242) = -.07$, $p = .290$; arousal: $r(242) = .65$, $p < .001$; valence: $r(242) = .50$, $p < .001$).

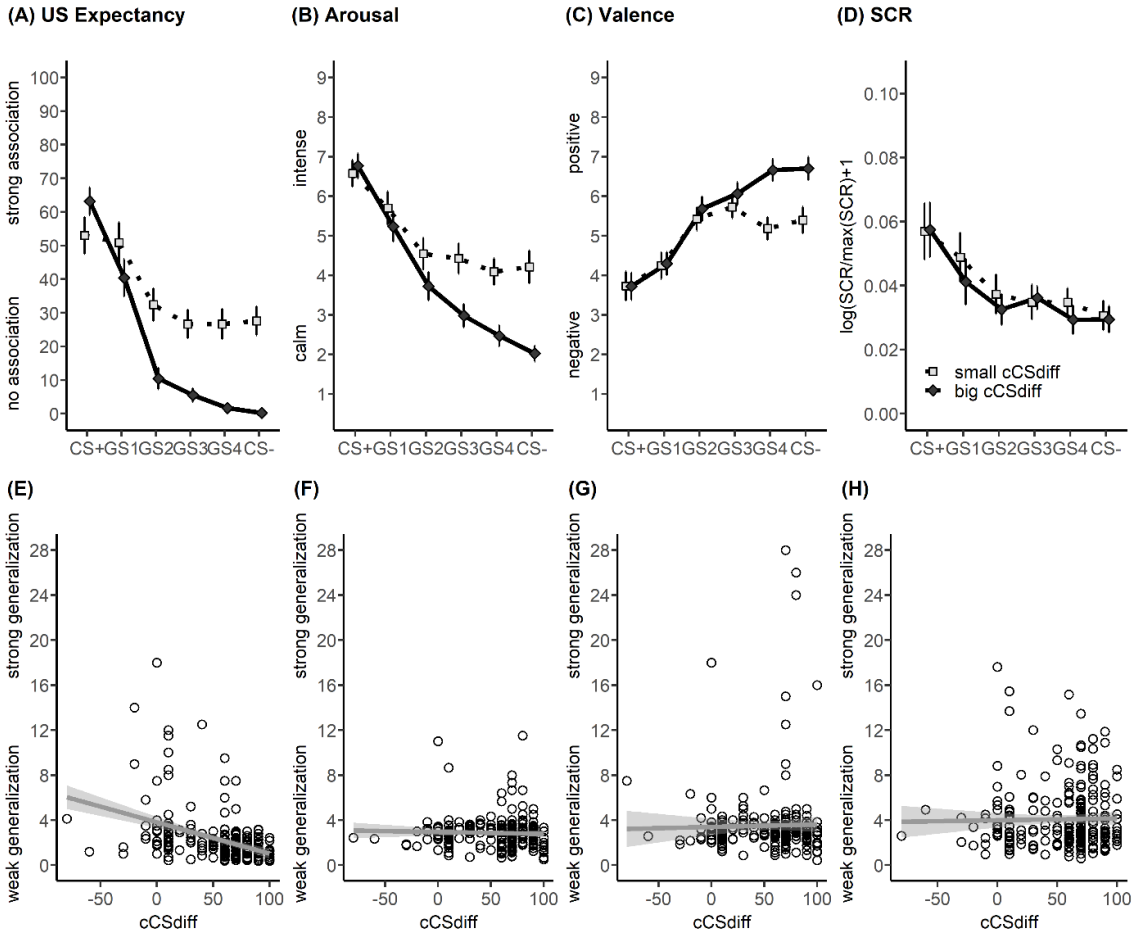


Figure 19. Pre-training generalization gradients and GI as a function of CS differentiation on cognitive level (cCSdiff).

Notes. The gradients show means (with SEs) for ratings of US expectancy (A), arousal (B), and valence (C), and skin conductance responses (SCR; D) of extreme groups with small and big cCSdiff, i.e., with a cCSdiff that is more than one SD below or above the mean, respectively. The respective GIs are plotted below (E-H). cCSdiff × Stimulus interaction was significant for US expectancy (A) and arousal ratings (B), post-hoc correlations of cCSdiff with GI were significant for US expectancy only (E).

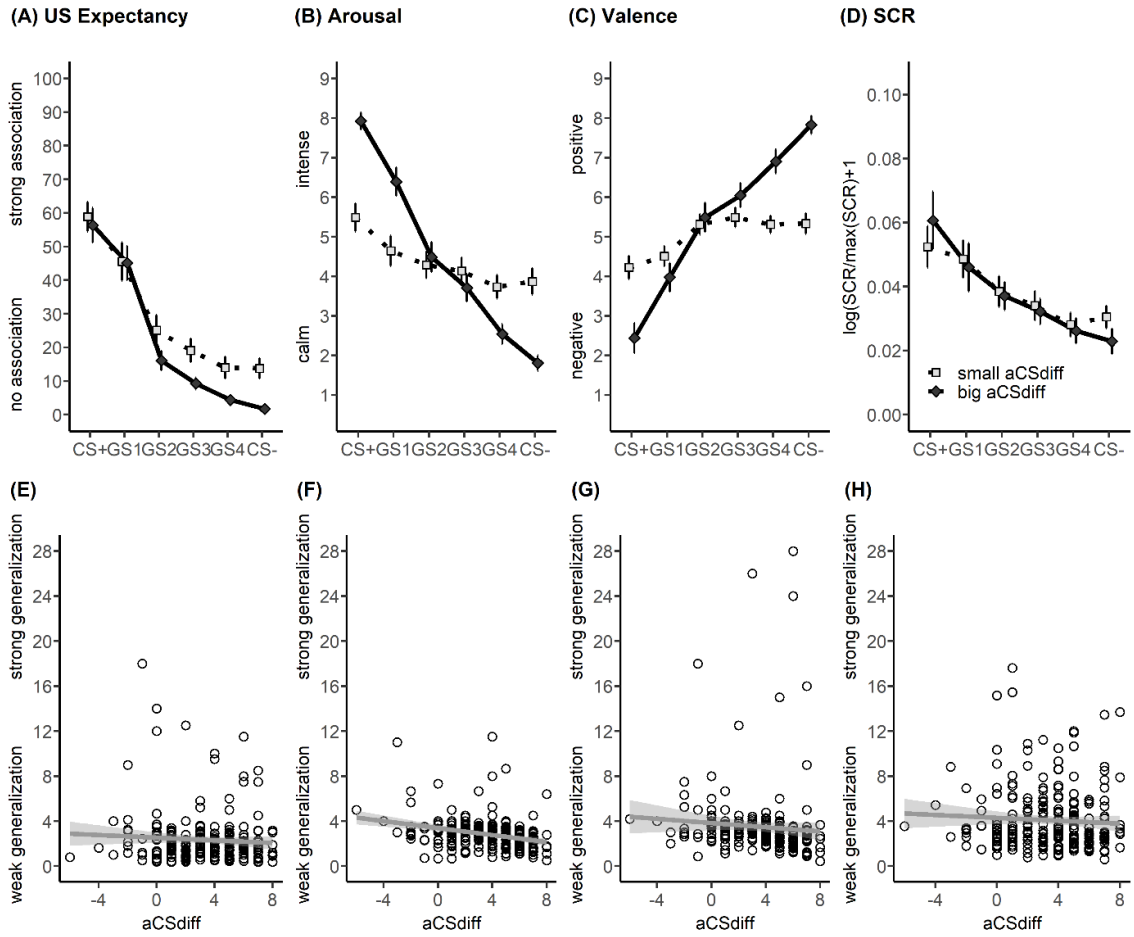


Figure 20. Pre-training generalization gradients and GI as a function of CS differentiation on affective level (aCSdiff).

Notes. The gradients show means (with SEs) for ratings of US expectancy (A), arousal (B), and valence (C), and skin conductance responses (SCR; D) of extreme groups with small and big aCSdiff, i.e., with an aCSdiff that is more than one SD below or above the mean, respectively. The respective GIs are plotted below (E-H). aCSdiff × Stimulus interaction was significant for all ratings (A-C), but not SCR (D). Post-hoc correlations of aCSdiff with GI were significant for arousal ratings only (F).

3.3.2.3 Impact of anxious personality on discrimination-training effects

The addition of the risk factor anxiety to analyses of discrimination-training effects explained no additional variance for neither of the outcome measures (for R^2 of the models see **Table 6**), as indicated by non-significant model comparisons (all *p*-values > .122).

Table 6 *Explained variance (R^2) of regression models*

	US expectancy	arousal	valence	SCR
Model A				
R^2	0.24	0.14	0.06	0.06
Model B (anxiety)				
R^2	0.25	0.15	0.08	0.09
Model C (cCSdiff)				
R^2	0.27	0.18	0.07	0.07
Model D (aCSdiff)				
R^2	0.28	0.23	0.10	0.07

Notes. cCSdiff or *aCSdiff* CS differentiation at the end of fear acquisition on cognitive level or affective level.

3.3.2.4 Impact of impaired fear vs. safety learning on discrimination-training effects

Significant comparisons of model A and model C for US-expectancy ($\chi^2(4, 233) = 2.65$, $p = .034$) and arousal ratings ($\chi^2(4, 235) = 2.84$, $p = .025$) indicated that the risk factor cognitive CS differentiation (cCSdiff) explained additional variance of discrimination-training effects. Conversely, model C was not superior for valence ratings and SCR (all p-values > .434). R^2 of the models are summarized in **Table 6**. Post-training generalization gradients and GI as a function of cCSdiff are shown for all outcome measures in **Figure 21** A-D and **Figure 21** E-H, respectively.

A look at the individual coefficients of model C for US-expectancy and arousal ratings (see **Table 7** and **Table 8**) showed a significant main effect of cCSdiff (US expectancy: $t(233) = 2.79$, $p = .006$; arousal: $t(235) = 3.01$, $p = .003$), indicating a negative association with generalization post training (see **Figure 21** E and **Figure 21**

F, respectively). The model revealed no significant interaction between cCSdiff and the training conditions (i.e., fear relevance and/or feedback) (all *p*-values > .127).

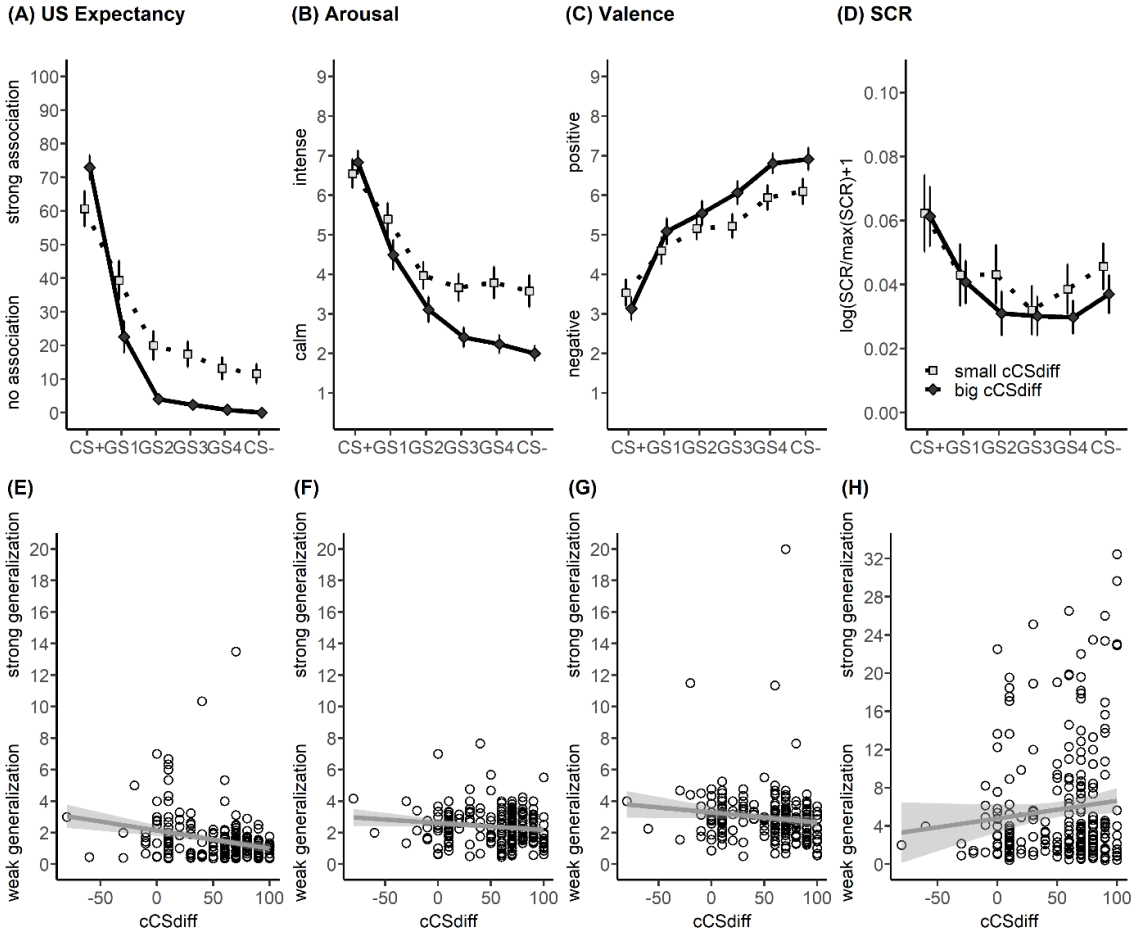


Figure 21. Post-training generalization gradients and GI as a function of CS differentiation on cognitive level (cCSdiff).

Notes. The gradients show means (with SEs) for ratings of US expectancy (A), arousal (B), and valence (C), and skin conductance responses (SCR; D) of extreme groups with small and big cCSdiff, i.e., with a cCSdiff that is more than one SD below or above the mean, respectively. The respective GIs are plotted below (E-H). Negative associations of cCSdiff and GI were found for US-expectancy (E) and arousal ratings (F).

Table 7 *Coefficients of Model C (including cCSdiff) for US-expectancy ratings*

	β	***SE***	***t***	***p***
GI_pre	0.47	0.06	7.65	**< .001**
relevance	0.23	0.15	1.51	.132
feedback	0.32	0.15	2.15	**.032**
Relevance × Feedback	0.10	0.15	0.66	.510
cCSdiff	-0.21	0.08	2.79	**.006**
cCSdiff × Relevance	-0.21	0.14	1.53	.127
cCSdiff × Feedback	-0.11	0.13	0.81	.420
cCSdiff × Relevance × Feedback	< -0.01	0.14	0.03	.977

Notes. GI generalization index, *cCSdiff* CS differentiation at the end of fear acquisition on cognitive level.

Table 8 *Coefficients of Model C (including cCSdiff) for arousal ratings*

	β	***SE***	***t***	***p***
GI_pre	0.35	0.06	5.83	**< .001**
relevance	< 0.01	0.16	0.03	.980
feedback	0.31	0.16	1.94	.054
Relevance × Feedback	-0.06	0.16	0.36	.716
cCSdiff	-0.24	0.08	3.01	**.003**
cCSdiff × Relevance	-0.09	0.14	0.62	.538
cCSdiff × Feedback	-0.14	0.14	1.02	.309
cCSdiff × Relevance × Feedback	0.10	0.14	0.67	.505

Notes. GI generalization index, *cCSdiff* CS differentiation at the end of fear acquisition on cognitive level.

Comparisons of model A and model D confirmed that the addition of risk factor affective CS differentiation (aCSdiff) to analyses of discrimination-training effects explained additional variance for US-expectancy ($\chi^2(4, 233) = 3.12$, $p = .016$), arousal ($\chi^2(4, 235) = 6.65$, $p < .001$) and valence ratings ($\chi^2(4, 233) = 2.64$, $p = .035$). The model comparison for SCR did not reach significance ($\chi^2(4, 235) = 0.7$, $p = .596$). R^2

of the models are summarized in **Table 6**. Post-training generalization gradients and GI as a function of aCSdiff are shown for all outcome measures in **Figure 22** A-D and **Figure 22** E-H, respectively.

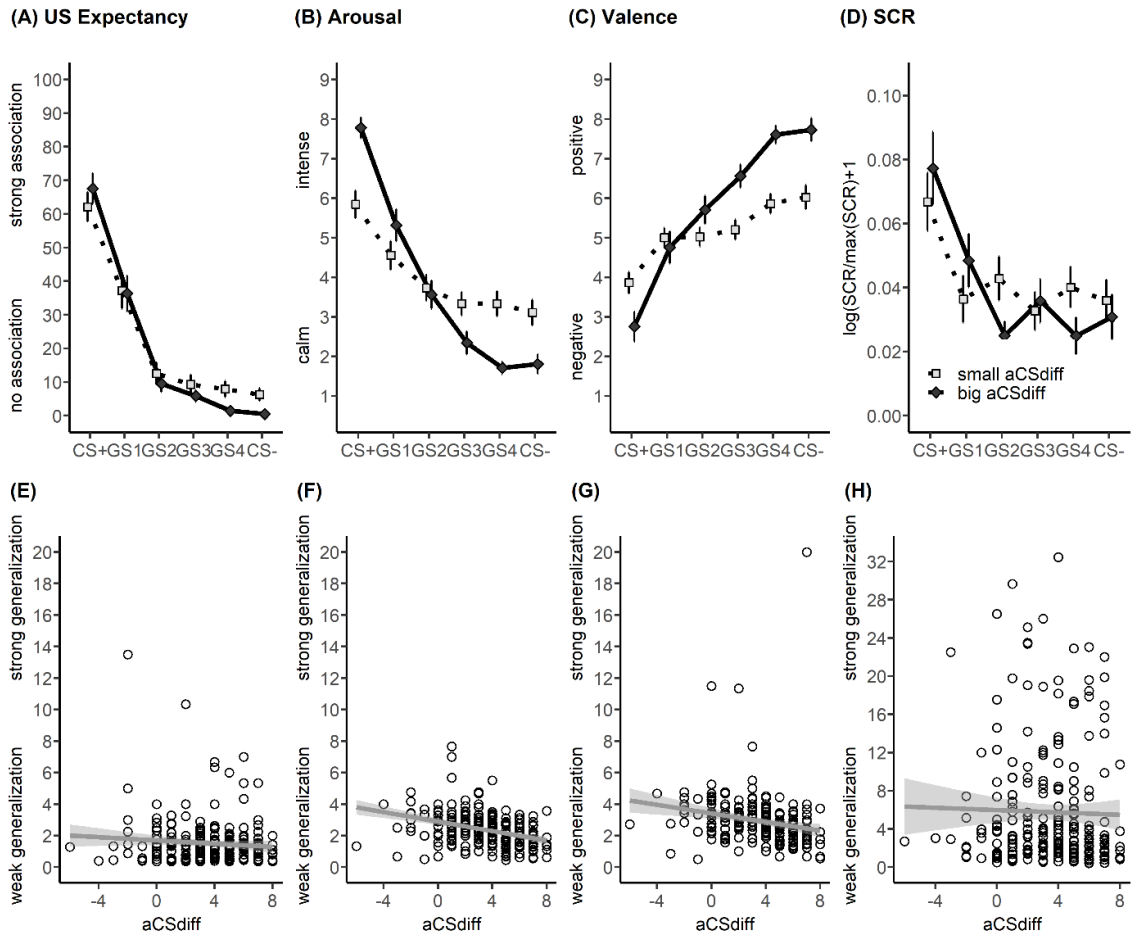


Figure 22. Post-training generalization gradients and GI as a function of CS differentiation on affective level (aCSdiff).

Notes. The gradients show means (with SEs) for ratings of US expectancy (A), arousal (B), and valence (C), and skin conductance responses (SCR; D) of extreme groups with small and big aCSdiff, i.e., with an aCSdiff that is more than one SD below or above the mean, respectively. The respective GIs are plotted below (E-H). Negative associations of aCSdiff and GI were found for arousal (F) and valence ratings (G).

A look at the individual coefficients of model D for all ratings (see **Table 9**, **Table 10** and **Table 11**) showed a significant main effect of aCSdiff for arousal ($t(235) = 4.35$, $p < .001$) and valence ratings ($t(233) = 3.10$, $p = .002$), indicating a negative association with generalization post training (see **Figure 22** F and **Figure 22** G, respectively). Instead of a main effect (see **Figure 22** A and **Figure 22** E), for US

expectancy, the model revealed a significant interaction of aCSdiff and the training condition feedback ($t(233) = 2.24$, $p = .026$; **Figure 23**).

Table 9 *Coefficients of Model D (including aCSdiff) for US-expectancy ratings*

	β	***SE***	***t***	***p***
GI_pre	0.51	0.06	8.72	**< .001**
relevance	0.14	0.09	1.53	.128
feedback	0.29	0.09	3.10	**.002**
Relevance × Feedback	0.13	0.09	1.44	.150
aCSdiff	-0.10	0.06	1.83	.068
aCSdiff × Relevance	-0.12	0.09	1.34	.183
aCSdiff × Feedback	-0.21	0.09	2.24	**.023**
aCSdiff × Relevance × Feedback	-0.13	0.09	0.03	.153

Notes. GI generalization index, a*CSdiff* CS differentiation at the end of fear acquisition on affective level.

Table 10 *Coefficients of Model D (including aCSdiff) for arousal ratings*

	β	***SE***	***t***	***p***
GI_pre	0.26	0.06	4.56	**< .001**
relevance	0.11	0.09	1.17	.244
feedback	0.02	0.09	0.22	.825
Relevance × Feedback	0.02	0.09	0.20	.843
aCSdiff	-0.27	0.06	4.35	**<.001**
aCSdiff × Relevance	-0.20	0.10	2.08	**.038**
aCSdiff × Feedback	0.05	0.10	0.53	.595
aCSdiff × Relevance × Feedback	-0.05	0.10	0.54	.588

Notes. GI generalization index, a*CSdiff* CS differentiation at the end of fear acquisition on affective level.

Table 11 *Coefficients of Model D (including aCSdiff) for valence ratings*

	β	***SE***	***t***	***p***
GI_pre	0.17	0.06	2.71	**.007**
relevance	0.06	0.10	0.62	.533
feedback	-0.14	0.10	1.35	.178
Relevance × Feedback	0.07	0.10	0.70	.487
aCSdiff	-0.20	0.06	3.10	**.002**
aCSdiff × Relevance	-0.06	0.10	0.54	.588
aCSdiff × Feedback	-0.02	0.10	0.12	.881
aCSdiff × Relevance × Feedback	-0.04	0.10	0.34	.735

Notes. GI generalization index, a*CSdiff* CS differentiation at the end of fear acquisition on affective level.

Additionally, for arousal ratings, the interaction of aCSdiff and the training condition fear relevance ($t(235) = 2.08$, $p = .038$) returned significant (**Figure 24**). All further effects of model D including aCSdiff did not reach significance (all p-values > .068). Post-hoc analysis of the aCSdiff × Feedback interaction revealed a positive association of generalization pre and post training in both feedback ($\beta = 0.56$, $SE = 0.07$, $t(125) = 7.51$, $p < .001$) and no-feedback groups ($\beta = 0.38$, $SE = 0.08$, $t(111) = 4.46$, $p < .001$). Strikingly, a negative association of aCSdiff and generalization post training was not observed in the feedback condition ($\beta = 0.02$, $SE = 0.07$, $t(125) = 0.34$, $p = .738$; **Figure 23** C, for generalization gradients of the respective aCSdiff extreme groups see **Figure 23** A), but in the no-feedback condition ($\beta = -0.24$, $SE = 0.08$, $t(111) = 2.88$, $p = .005$; **Figure 23** D, for generalization gradients of the respective aCSdiff extreme groups see **Figure 23** B). Post-hoc analysis of the aCSdiff × Fear relevance interaction revealed a positive association of generalization pre and post training in both fear-relevant ($\beta = 0.31$, $SE = 0.09$, $t(116) = 3.43$, $p < .001$) and fear-irrelevant training groups ($\beta = 0.25$, $SE = 0.08$, $t(122) = 3.15$, $p = .002$). Importantly, a negative

association of aCSdiff and generalization post training was not observed in fear-relevant training groups (β = -0.15, SE = 0.09, $t(116) = 1.63$, $p = .105$; **Figure 24** C, for generalization gradients of the respective aCSdiff extreme groups see **Figure 24** A), but in fear-irrelevant groups only (β = -0.40, SE = 0.08, $t(122) = 4.98$, $p < .001$; **Figure 24** D, for generalization gradients of the respective aCSdiff extreme groups see **Figure 24** B).

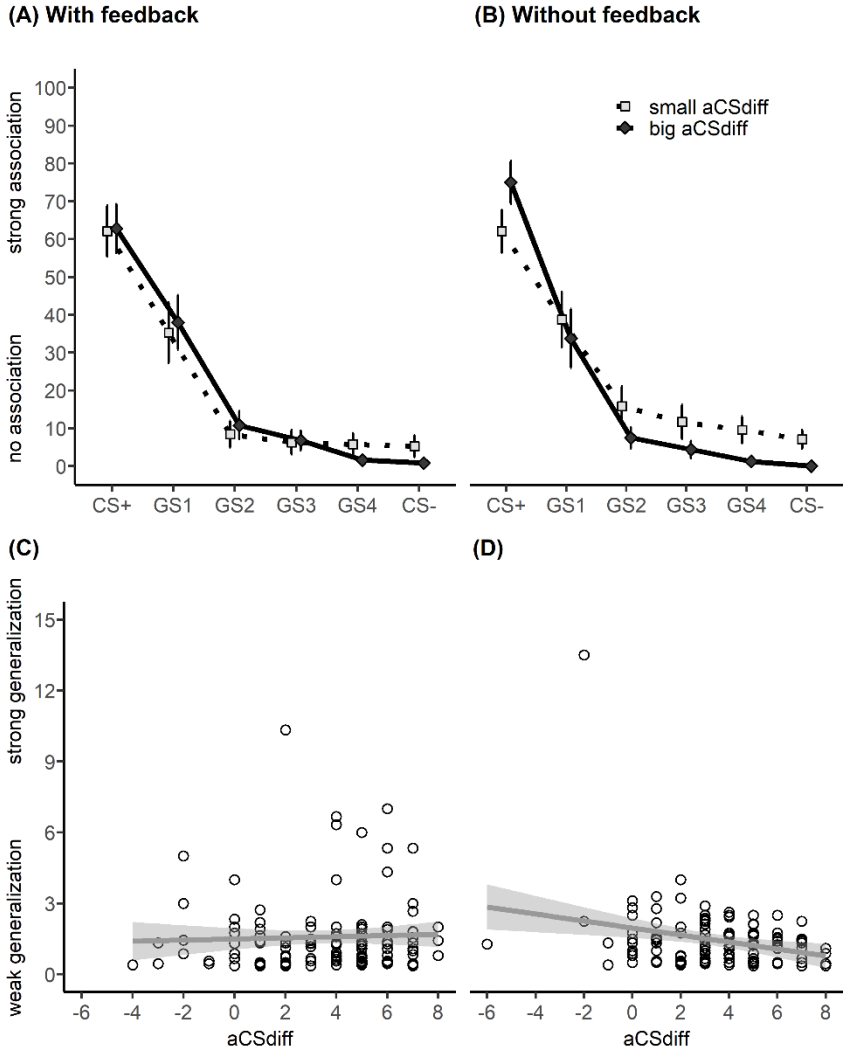


Figure 23. Post-training generalization of US expectancy in feedback vs. no-feedback groups as a function of CS differentiation at the affective level (aCSdiff).

Notes. The upper panels show gradients with means (with SEs) of extreme groups with small and big aCSdiff, i.e., with a aCSdiff that is more than one SD below or above the mean, in groups with feedback (A) and without feedback (B), respectively. A negative association of aCSdiff and GI was not observed in the feedback condition (C), but in the no-feedback condition (D).

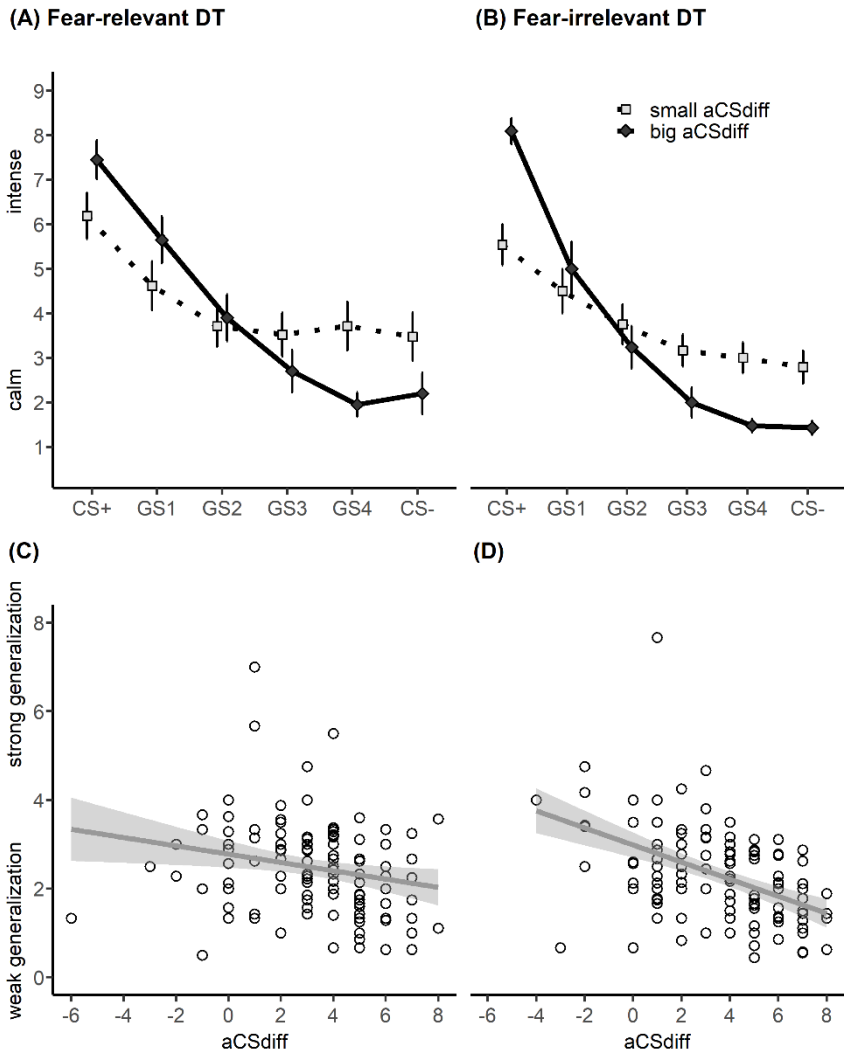


Figure 24. Post-training generalization of arousal in fear-relevant vs. fear-irrelevant discrimination training groups as a function of CS differentiation at the affective level (aCSdiff).

Notes. The upper panels show gradients with means (with SEs) of extreme groups with small and big aCSdiff, i.e., with a aCSdiff that is more than one SD below or above the mean, in fear-relevant (A) and fear-irrelevant (B) training groups, respectively. A negative association of aCSdiff and GI was not observed in the fear-relevant DT condition (C), but in the fear-irrelevant DT condition (D).

3.3.3 Discussion

Study 2B addressed the question how risk factors of anxiety disorder influence fear generalization and its reduction by discrimination trainings. Specifically, I selected the vulnerability factors trait anxiety and deficient learning of threat-safety differentiation (CS differentiation, Mineka & Zinbarg, 2006). The latter was additionally distinguished in learning on cognitive and affective level as reflected in

US-expectancy and arousal ratings, respectively (Kleinginna & Kleinginna, 1981; Lonsdorf et al., 2017).

The first hypothesis was that fear generalization becomes more pronounced with increasing degrees of anxiety. Contrary to expectations, results showed that trait anxiety does not influence the generalization gradients' steepness but is associated with a higher reaction level to all stimuli during generalization test pre training. Correspondingly, no association of trait anxiety and generalization (as reflected in generalization index GI) was found post training and no differences in training responsiveness depending on trait anxiety could be revealed. In consequence, my analyses were not able to answer the further hypothesis that high trait anxious individuals may particularly benefit from fear-relevant discrimination training or training with feedback. At first glance, the finding that higher anxiety leads to a stronger response to all stimuli contradicts previous findings that found a positive association of anxiety and generalization, more precisely the linearity of the gradient (Baumann et al., 2017). This is even more surprising since I used the same paradigm and questionnaires as Baumann et al. (2017). However, it must be taken into account that the aforementioned study had a comparatively small sample and determined effects in extreme group comparison. Based on the present results, it is only logical that the post-training GI analysis showed no effects of anxiety. After all, the GI indicates the relationship between the individual stimuli and not the response level. The result fits the general picture painted by the literature. Studies predominantly reported small effects of trait anxiety on fear generalization in the sense of a flat gradient (Baumann et al., 2017; Haddad et al., 2012; for a meta-analysis see Sep et al., 2019) or contradictory results (e.g., Torrents-Rodas et al., 2013). It is possible that the positive association of anxiety and fear generalization occurs only in very highly anxious individuals, which could be an explanation for the mixed findings (Andreatta & Pauli, 2017). Considering that the current study had very strict inclusion criteria,

resulting in a very healthy sample, this could also explain the lacking effect of anxiety on GI here. However, on average, the sample already reaches the STAI value reported by Torrents-Rodas et al. (2013) for the highly anxious subsample, suggesting that its anxiety level was not particularly low (see **Supplementary Table 7**). However, it should be noted that in the current study, individual anxiety is not based on the STAI, but on the Baumann anxiety factor (Baumann et al., 2017), to whose anxiety level the present study is comparable.

Two strengths of the current study contribute to the validity of the finding that anxiety is associated with stronger response levels to all stimuli. First, the analyses were conducted on a large sample, which allows for reliable results with good power (Button et al., 2013; Case & Ambrosius, 2007; Mendoza, Stafford, & Stauffer, 2000). Second, to my knowledge, it is the first study to examine the relationship between fear generalization and anxiety, using the latter as a continuous variable. In contrast, previous studies have mostly used extreme group designs where effects are easier to find (Fisher, Medaglia, & Jeronimus, 2018; Preacher, Rucker, MacCallum, & Nicewander, 2005), which may have led to an overestimation of the true effect. I would therefore assume that anxiety can affect both the shape of the gradient and general reactivity to stimuli.

With respect to varying degrees of CS differentiation after acquisition phase, I hypothesized that fear generalization would be more pronounced with increasing individual vulnerability. Consistent with the hypothesis, individual differences in the threat-safety differentiation at both the cognitive (cCSdiff) and affective (aCSdiff) levels affected the shape of the generalization gradients before training, particularly the response to the CSs. Consistently, negative associations of cCSdiff and aCSdiff with generalization were also observed after training. However, it should be noted that both

findings were limited to ratings. In the post training generalization test, I expected persistently higher fear generalization (i.e., lower response to training) in more susceptible individuals who trained with fear-irrelevant stimuli or without feedback. Consistent with this hypothesis, I found a negative association of aCSdiff and generalization of US expectancy in groups that did not receive feedback during discrimination training. Similarly, I found a negative association between aCSdiff and the generalization of arousal in groups that trained with fear-irrelevant stimuli.

As noted above, a key finding is that participants with greater CS differentiation during learning also show greater differentiation during generalization (in agreement with Stegmann et al., 2019). Greater CS differentiation affects the generalization gradient by increasing the range between minimum and maximum response of the gradient, similar to an anchor effect (Brown, 1953). Thus, we should not underestimate the predictive value of CSdiff, especially because poor CS differentiation can be interpreted as generalization (Lissek et al., 2009). While CS differentiation was significantly correlated across phases, a significant correlation between CSdiff and generalization index (GI) was only demonstrated within the same outcome measure. One explanation for the latter correlations not being significant despite Stimulus × CSdiff interaction is that the GI does not contain CS-, but the interaction is likely driven by CS-difference (for which CS- is essential). Second, the GI reflects the relationship between stimuli within an individual, and individual generalization patterns (GIs) may differ despite equal CSdiff, as in the case of generalization clusters 2 and 4 in the study by Stegmann et al. (2019). In contrast, the ANCOVA interaction indicates that the average response of the sample to single stimuli varies according to individual CSdiff. One possible conclusion is that individual generalization (as reflected in GI) can be considered a separate risk factor as it seems to provide information beyond CSdiff.

Nonetheless, the significant correlation of GI and CSdiff within the same outcome measure suggest that GI is sensitive to CSdiff.

Remarkably, negative associations of cCSdiff and aCSdiff with generalization were also observed after training. Importantly, in the meantime, discrimination training was supposed to help participants better discriminate between stimuli. Moreover, the generalization test after training provided another opportunity to learn contingencies. Thus, one could conclude that the CSdiff risk factor has predictive value regarding the steepness of the gradient for individuals who have difficulty discriminating between stimuli despite the training and the second generalization block.

Of particular interest is the limited evidence that different training conditions appear to have different effects as vulnerability increases. This is provided by the negative association between aCSdiff and the generalization of US-expectancy ratings in groups that did not receive feedback during discrimination training. The same is suggested by the negative association between aCSdiff and the generalization of arousal in groups that trained with fear-irrelevant stimuli. A first look at the plots (**Figure 23** and **Figure 24**) suggests interpreting the interaction between aCSdiff and training conditions as follows. Individuals who showed poor CS differentiation at the affective level during learning particularly benefit from discrimination training with fear-relevant vs. fear-irrelevant stimuli or training with vs. without feedback. These training conditions seem to compensate for the differences in affective learning. In line, the generalization gradient plots (**Figure 23** A-B and **Figure 24** A-B) show more similar generalization gradients in both aCSdiff extreme groups in the training conditions with relevant stimuli and feedback, unlike in other training conditions. Furthermore, groups with low aCSdiff descriptively exhibit a steeper generalization gradient in the feedback vs. no feedback group or in the fear-relevant vs. fear-irrelevant group. However, the GI

plots (**Figure 23** C-D and **Figure 24** C-D) also reveal that the positive effect of fear relevance and feedback may be overestimated, as single data points may have strongly influenced the regression lines. Candidates for this might be, for example, a data point in the fear-relevant group suggesting that the polarity of CSs is confounded during acquisition but not during generalization, or another data point in the fear-irrelevant group and no-feedback group with comparatively high GI. Therefore, in general, the greater success of certain training conditions in at-risk individuals should be interpreted with great caution.

It is noticeable that, if anything, CSdiff at the affective level, but not at the cognitive level, leads to a different strength of response to training conditions. Also, cognitive and affective CSdiff have different effects on the generalization gradient independent of the training conditions. This is shown by the extreme group plots as cCSdiff affects response strength at the safety end of the gradient (see **Figure 21**), whereas aCSdiff affects response strength at both safety and threat ends of the gradient (see **Figure 22**). In interpreted terms, this dissociation of affective and cognitive levels means that contingency awareness does not preclude anxious affect. This is in line with the meta-analysis of W. Hofmann et al. (2010) that affective learning requires more time (i.e., trials) than cognitive learning of US contingencies.

3.4 Discussion Study 2

The current experiment was conducted to replicate the findings of the proof-of-principle study (Study 1A), that ascertained fear generalization can be successfully reduced by fear-relevant discrimination training or feedback on discrimination performance. In addition, I examined how risk factors for anxiety disorders influence fear generalization and its reduction through discrimination training, as affected individuals have a particular need for fear-reducing interventions. Specifically, I

hypothesized there would be an increase in fear generalization with increasing individual vulnerability, i.e., higher anxiety and lower CS differentiation. Furthermore, I expected that fear relevance and feedback would compensate for vulnerability and consequently lead to good responsiveness in general, whereas responsiveness to fear-irrelevant discrimination training or training without feedback would be lower as vulnerability increases.

To investigate these research questions, a large sample of healthy participants underwent the experimental paradigm established in Study 1A and were asked about their anxiety characteristics.

As expected, feedback improved discrimination performance in training right from the beginning (Dosher & Lu, 2017; M. H. Herzog & Fahle, 1999). Specifically, it facilitated discrimination learning for particularly similar stimuli (CS+ and GS1 or the two thinnest lines). In addition, the feedback probably invited the inclusion of higher-order processes to maximize performance (e.g., rule learning; Goodman, Wood, & Chen, 2011). Rewarding the detection of very small differences particularly invites to infer the rule that discrimination of similar stimuli is purposeful in this experiment. The sustained action of these perceptual and learning processes in subsequent generalization testing results in a particularly effective reduction of fear generalization compared to training without feedback. In addition to US-expectancy rating, this time the feedback effect was also observed for physiological data (SCR). The credibility of this feedback effect increases further with confirmation that it was not driven by limiting group differences before training (as discussed in Study 2A, section 3.2.3), as the effect for US-expectancy ratings persisted even when controlling for CS differentiation at the end of learning at the cognitive level (cCSdiff), while the contradictory feedback effect for

valence ratings disappeared when controlling for CS differentiation at the end of learning at the affective level (aCSdiff; see Study 2B, section 3.3.2.4).

However, in contrast to the previous study, no support was found for stronger effects of fear-relevant vs. fear-irrelevant discrimination training. As previously reported, discrimination training even with fear-irrelevant stimuli can generally reduce subsequent generalization effects (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017), presumably because multiple processes, e.g. attentional and rule learning processes, are involved in perceptual learning (Dosher & Lu, 2017) and contribute to fear reduction in the subsequent generalization test. In contrast, perceptual learning itself is usually very specific and cannot easily be transferred to other stimulus material (Dosher & Lu, 2017; Furmanski & Engel, 2000). That fear relevance of the stimuli during discrimination training made no difference in the here examined healthy sample raises the question of the extent to which fear generalization and effects of discrimination-training are due to perceptual discrimination. It seems that participants were already good at perceptually discriminating CS+ from the other facial stimuli before training (for a broader discussion see section 4.2) because all of them (especially CS+) had been shown many times before and the similarity graduations of 20% steps were not too small. Consequently, perceptual processes and thus also fear relevance of the training stimuli did not play a role in fear reduction in the healthy sample. Instead, higher-order processes that were equally addressed in fear-relevant and fear-irrelevant discrimination training were transferred to post-training generalization, resulting in comparable levels of fear generalization.

For at-risk individuals (e.g., with low CS difference at the affective level), the situation might be different, as the current study revealed that fear relevance of the training stimuli or feedback on training performance might compensate for the

weaknesses in affective learning. The inclusion of the risk factor as a continuous variable revealed that the success of discrimination training depends on the expression of risk only in the conditions with fear-irrelevant stimuli and without feedback. Compared to cognitive learning, affective learning is more difficult to achieve (i.e., needs more trials; W. Hofmann et al., 2010; Vansteenwegen et al., 2006). This is where feedback and relevance, which facilitate new learning and its transfer to similar other tasks (Dosher & Lu, 2017; Furmanski & Engel, 2000; Sasaki et al., 2010), have a beneficial effect. This knowledge could be helpful in selecting a promising training, in the spirit of more individualized therapy (S. G. Hofmann & Hayes, 2018). However, since these small effects were likely favored by single data points, they should be interpreted with great caution. In addition, I recommend that future studies more specifically examine the influence of vulnerability on the effect of different training conditions, for example, by experimentally manipulating vulnerability (i.e. contingency awareness, for an example see Tabbert, Stark, Kirsch, & Vaitl, 2006).

Regardless of the training condition, the current study showed that the risk factor low CSdiff on cognitive and affective levels predicted more generalization post training despite training and a second generalization block. Pre training the same relation was found within the same outcome measure only, emphasizing that learning on cognitive and affective level initially are distinct processes (LeDoux & Pine, 2016; Ohman & Mineka, 2001). The more interesting is the finding that poor learning at either level becomes a long-term risk factor for increased fear generalization across all outcome measures. Therefore, my results once again suggest that poor CS differentiation can be interpreted as generalization (Lissek et al., 2009), and consequently, we should not underestimate the predictive value of CSdiff.

In contrast, trait anxiety did not affect the shape of the generalization gradient but increased the response level for all ratings, likely due to a heightened perception of arousal to fearful stimuli (Lang & McTeague, 2009; Rosebrock, Hoxha, Norris, Cacioppo, & Gollan, 2017). The reported experience of increased arousal may reflect cognitive and attentional biases (Fisher, Granger, & Newman, 2010). The result is consistent with previous findings on fear generalization that anxiety is associated with a higher response level rather than a narrow slope (Stegmann et al., 2019). Data from other studies also suggest higher response levels in high anxious individuals, but the main effects of anxiety have not been reported (Baumann et al., 2017; Mertens, Bouwman, & Engelhard, 2021; Torrents-Rodas et al., 2013). The current study extends the validity of this finding to anxiety used as a continuous variable. In future, therefore, it would be interesting to investigate whether discrimination training can also reduce response levels. Of course, this conclusion raises the question of whether the here used training, which aims at fear reduction in the sense of steepening the generalization gradient, is a useful approach for anxious individuals or even anxious patients. However, as very high levels of anxiety (e.g., anxiety patients) might be associated with a flatter curve of the generalization gradient (Andreatta & Pauli, 2017), training might be particularly useful for this group of individuals.

Taken together, the results on the influence of risk factors on the generalization of fear suggest that both trait anxiety and deficient CS differentiation influence fear responses to generalization stimuli. Importantly, the relation was examined on individual level, emphasizing the view of risk factors as an individual trait, in line with conceptualization of psychopathology as a continuum (Cuthbert, 2014; Insel et al., 2010). Nevertheless, fear generalization can be considered a distinct risk factor because it is not fully predetermined by the other two. As the lack of association of anxiety or CS difference (beyond the same outcome measure) with fear generalization pre training

illustrate, the message of fear generalization goes beyond reaction level or maximum differentiability between fear and safety. Consequently, all three risk factors encompass different aspects of fear and anxiety. Some studies already tried to predict clinical anxiety with fear acquisition, showing mixed results (for a review see Scheveneels, Boddez, & Hermans, 2021). However, to date, there is only one published study examining whether individual differences in generalization of fear can predict the level of anxiety at a later point in time (Lenaert et al., 2014). Indeed, the study found evidence that broader generalization to stimuli more similar to the safety stimulus than to the threat stimulus predicts higher anxiety 6 months later. Further research on this topic is urgently warranted to determine whether fear generalization is indeed a risk factor with predictive power for clinical anxiety.

This study has important strengths, e.g., large sample size and analysis of individual differences, but also some weaknesses. For example, there were group differences prior to training. Therefore, the feedback effect for US-expectancy ratings may have been favored by the higher initial values, as they have a greater potential for reduction (Jennings & Cribbie, 2016). On the other hand, the main effect of feedback remains when controlling for differences between fear vs. safety learning. The latter have substantial influence on generalization scores pre-training. Since the effect remains, the pre-training differences seem to have added only to the true effect. Moreover, taking into account the differences in fear vs. safety learning, the contradictory feedback effect for valence ratings is fortunately absent.

As in the first study, the effects of the between factors were again mainly limited to ratings. Although in this study the fear-reducing effect of training with feedback was also demonstrated for SCR, it is difficult to interpret because, as in Study 1, all training conditions increased GI of SCR. Cognitive and affective CS difference risk factors did

not affect fear generalization for SCR neither before nor after training. A possible reason for this would be that the SCR reactions have been standardized, so that individual (e.g., level) differences are not taken into account. Moreover, it makes sense that three output systems of emotions are distinguished, namely physiological, subjective verbal and behavioral (Bradley & Lang, 2000; Lang, 1968), because their output can differ (e.g., in the degree of generalization, Holt et al., 2014; Schiele et al., 2016). In line, in this study, a more pronounced generalization of fear pre training was found at the rating level than at the physiological level. Moreover, the risk factor CS differentiation at both cognitive and affective levels correlated with the generalization index mostly only within the same outcome measure (i.e., US-expectancy and arousal ratings) and never with physiological data (i.e., SCR).

In summary, the current study provided further evidence for the successful reduction of fear generalization by my "therapeutic" approach of discrimination training, whereby training is implemented only after fear acquisition and an initial demonstration of fear generalization. Specifically, implementing reinforcing feedback during training reduced fear generalization particularly well, for at least cognitive fear parameters, i.e., US expectancy, as feedback is likely to directly increase attention and motivation for perceptual tasks. In contrast, fear relevance of discrimination training did not prove to be particularly fear-reducing in healthy participants. However, the study found limited evidence that individuals at risk for an anxiety disorder (i.e., deficient in fear vs. safety learning) might benefit from the fear relevance of the training. In addition, the study showed that the risk factors trait anxiety and lack of fear vs. safety learning relevantly but not completely predicted different aspects of fear generalization independent of training conditions, so the latter can still be considered a distinct risk factor. Importantly, in contrast to most previous research, this study accounted for individual differences in anxiety disorder risk at a continuous level, consistent with the

conceptualization of psychopathology as a continuum (Cuthbert, 2014; Insel et al., 2010). Accordingly, this is also the first attempt to predict the success of discrimination training based on the severity of risk, adhering to the idea of more individualized therapy (S. G. Hofmann & Hayes, 2019).

4 General Discussion

The purpose of this dissertation was to examine, for the first time, whether discrimination training can reduce the extent of fear generalization following fear acquisition and an initial demonstration of generalization. Because patients typically seek therapy at this time, this approach mimics a therapeutic intervention. Improving stimulus discrimination is a promising approach, as fear generalization to similar stimuli appears to be related to the inability to perceptually discriminate between them (Holt et al., 2014; Zaman et al., 2019).

In Study 1, I examined fear-reducing effects of fear relevance and feedback in discrimination training and tested efficacy relative to non-discriminative control training. Indeed, I found evidence for a somewhat more effective reduction of ascertained fear generalization in healthy individuals when discrimination training was conducted with fear-relevant stimuli compared to fear-irrelevant stimuli, at least for cognitive fear parameters, i.e., US expectancy. The small effect could not be explained by the fact that all participants were trained in discrimination, because neither discrimination training reduced fear more than a non-discriminative control training. Therefore, other (higher-level) processes besides perceptual discrimination are likely to be involved in both discrimination and control training and contribute to fear reduction. Moreover, the study revealed that reinforcing feedback during discrimination training reduces generalization of US expectancy. In contrast, the feedback condition was no more successful than the no-feedback condition during non-discriminative control training. Consequently, feedback presumably works via motivational mechanisms that specifically increase commitment to perceptual tasks.

Study 2 served as a replication study of Study 1 and additionally examined how risk factors for anxiety disorders, namely higher anxiety and poorer ability to

discriminate between threat and safety stimuli, influence fear generalization and its reduction through discrimination training, as affected individuals have a particular need for fear-reducing interventions. As in Study 1, implementing reinforcing feedback during training reduced fear generalization particularly well, especially for US-expectancy ratings. However, unlike the previous study, no support was found for fear relevance of discrimination training being particularly beneficial for reducing fear generalization in healthy participants. It appears that other processes are involved in perceptual learning (Dosher & Lu, 2017) that promote fear reduction following discrimination training regardless of fear relevance. Yet, the study found limited evidence that fear relevance and feedback can compensate for vulnerability (i.e., poor fear learning compared to safety at the affective level, i.e., arousal ratings), resulting in good responsiveness and fear reduction as strong as in low-vulnerability individuals. In addition, the study showed that the risk factors trait anxiety and lack of fear vs. safety learning relevantly, but not completely, predicted different aspects of fear generalization independent of training conditions. Lower rates of learning CS differentiation at the cognitive and affective levels were predictive of the steepness of subjects' generalization gradient. In contrast, anxiety increased the response level for all ratings. It remains to be clarified whether very high levels of anxiety (e.g., anxiety patients) are associated with a flatter generalization gradient curve (as discussed by Andreatta & Pauli, 2017). Affected individuals could then benefit from my discrimination training.

4.1 Distinguishing different levels of learning and expressing the fear response

As noted in the summary of key findings in the previous section, the studies presented here collected and distinguish ratings at the cognitive and affective levels. As suggested earlier (Lonsdorf et al., 2017), the verbal level of emotional response reflects

the outcome of these two learning processes. Learning of CS-US contingencies (Rescorla, 1968) is a cognitive learning process reflected in US-expectancy ratings. In comparison, affective learning processes underlie the change in emotional impact of the stimuli experience (Hermans et al., 2002) reflected in valence and arousal ratings (Lonsdorf et al., 2017). Besides, emotional responses are expressed on physiological level, which can for instance be conducted by recording SCRs (Bradley & Lang, 1999; Lang, 1968; Lonsdorf et al., 2017). When several outcome measures indicating fear are included in fear conditioning research, they are often US-expectancy ratings for the verbal level and physiological measures (Beckers et al., 2013; Boddez et al., 2013; Lonsdorf et al., 2017) and are used for cross-validation. Accordingly, divergence between the measurements are undesirable (Beckers et al., 2013). Yet, it could contribute to the understanding of pathological fear (Beckers et al., 2013).

In this respect, it is a strength of the here presented studies that they all include ratings reflecting the cognitive learning level (i.e., US expectancy) and the affective learning level (i.e., arousal and valence) as well as a physiological measure (i.e., SCRs). While SCRs were recorded continuously throughout the experiment, ratings were assessed intermittently at the end of each block. Indeed, my findings revealed a dissociation between cognitive vs. affective learning. Furthermore, ratings differ from physiological measures. This is evident as less generalization was observed on physiological level (i.e., SCR) compared to verbal level (i.e., ratings). At rating level, generalization was less pronounced at cognitive level than at affective level.

In fact, it appears to be a robust phenomenon that fear generalization is broader at the rating level than at the physiological level (Haddad et al., 2013; Holt et al., 2014; Lissek et al., 2008; Torrents-Rodas et al., 2013). Such a conservative bias in verbal fear responses to a potential threat can activate attentional resources to gather more

information about the stimulus. Once this information indicates the presence of a threat, the autonomic fear system becomes active (Holt et al., 2014; LeDoux, 2000; LeDoux & Pine, 2016).

The divergence of cognitive and affective ratings is also not uncommon in the generalization literature (Ahrens et al., 2016; Meulders, Harvie, Lorimer Moseley, & Vlaeyen, 2015; Stegmann, Ahrens, Pauli, Keil, & Wieser, 2020), but is often not the subject of discussion. To this end, it must also be said that based on my review of the generalization literature, only some of the few studies that assessed both cognitive and affective ratings allow for a comparison of the respective extent of generalization. Yet there are legitimate reasons for divergence. For example, Boddez et al. (2013) emphasized that US-expectancy ratings are a valid index of fear learning, but that they cannot capture all aspects of fear and anxiety. After all, fear is probably more than the expectancy of an aversive outcome. For this reason, the authors recommend combining different measures, including self-report of arousal. Furthermore, the learning levels differ in sensitivity for new learning experiences. Obviously, new learning can be seen more easily and earlier on cognitive level, i.e., in US-expectancy ratings. In contrast, extinction after evaluative conditioning, reflected in changes in valence, is hard to achieve (W. Hofmann et al., 2010; Vansteenwegen et al., 2006).

Accordingly, in the studies presented here, learning levels also diverged with respect to the effects of discrimination training. First, the beneficial effects of the here presented discrimination training were mainly restricted to US-expectancy ratings. Second, there was limited evidence that fear relevance and feedback in training conferred benefits for individuals with poor threat vs. safety differentiation by the end of the acquisition phase at the affective level (reflected in arousal ratings, aCSdiff), but not the cognitive level (reflected in US-expectancy ratings, cCSdiff). Moreover, weak

cCSdiff or aCSdiff already pre-training affected the shape of the generalization gradient differently. This can be easily seen from the plots of the extreme groups (e.g., **Figure 19** and **Figure 20**), which show that both a better cCSdiff and aCSdiff lead to lower responses to stimuli at the safe end of the gradient. However, a better aCSdiff additionally leads to a stronger response to stimuli at the threatening end of the gradient, suggesting that participants with a poor aCSdiff also have difficulty to verbally express their subjective arousal towards CS+.

This suggests that affective ratings may provide more information about vulnerability to pronounced levels of fear or anxiety because they include uncertainty about stimuli despite contingency awareness. Accordingly, fearful individuals may well be aware of the irrationality of their exaggerated fear because they are cognitively aware that the feared object is most likely to have no aversive consequences. Nevertheless, they experience fear and exhibit fear behaviors.

From this, one can nicely see how the divergence of affective and cognitive ratings can contribute to the understanding of pathological fear (Beckers et al., 2013). It is a strength of this dissertation that the two learning processes were consistently assessed and even specifically distinguished in Study 2. Herein, I would like to follow previous recommendations (Beckers et al., 2013; Boddez et al., 2013; Taschereau-Dumouchel, Michel, Lau, Hofmann, & LeDoux, 2022) and encourage researchers studying human fear conditioning to include subjective measures of fear such as verbal self-report, distinguishing between cognitive and affective levels of learning.

Finally, a dissociation between outcome measures also indicates that likely higher-order processes are involved in fear generalization and its reduction. Accordingly, generalization does not only depend on perceptual similarity, but also on inferential reasoning processes which then can lead to gradients that deviate from

similarity-based gradients (Holt et al., 2014; Wong & Lovibond, 2017). This topic is discussed in more detail in the following section.

4.2 Perceptual discrimination and further processes involved in fear generalization and its reduction

The original idea of the discrimination training developed here was based on evidence that fear generalization also depends on the perceptual discriminability of very similar stimuli (Holt et al., 2014; Struyf et al., 2017; Zaman et al., 2019). Therefore, training was planned to reduce fear generalization by improving perceptual discrimination. Importantly, the results of this dissertation consistently show that multiple processes, including higher-level processes, appear to contribute to the fear-reducing training effects. This could be primarily because multiple mechanisms are involved in perceptual learning, such as sensory processing, decision-making and learning processes, attention, and feedback (Dosher & Lu, 2017). The activation of attentional processes like working memory was also promoted by trial structure, requiring the comparison of two subsequently presented stimuli (Lenaert et al., 2016).

This raises the question of the extent to which found discrimination-training effects are due to perceptual discrimination. For this, it is worth looking at discrimination performance during training (see Appendix 6.1.2 and 6.3.2). Both studies showed that participants improved perceptual discrimination of training stimuli, but not substantially. In fact, it appears that participants were already good at perceptual discrimination prior to training. In particular, this is suggested by the observation that even during fear-irrelevant training, when participants were unfamiliar with the stimuli, discrimination errors were only concentrated on the most similar stimuli and the reliable identification of identical stimuli. The latter observation also suggests that participants

expected to have to discriminate stimuli with greatest similarity. Moreover, feedback led to better performance during training when it was discriminative, whereas feedback had no effect on how participants solved a non-discriminative arithmetic task.

Consequently, other mechanisms were most likely addressed during training that subsequently reduced fear generalization. These may include processes of attentional control, i.e., working memory, that help focus attention on small perceptual differences, thereby facilitating comparison of stimuli and inhibition of responding to distractors (i.e., highly similar stimuli; Baddeley, 2012; Derakshan & Eysenck, 2009; Eysenck et al., 2007). Moreover, rules can be derived during discrimination training like paying attention to visual features of stimuli (Livesey & McLaren, 2009). Feedback also had a positive effect that was specifically investigated in this dissertation. Remarkably, this fear-reducing effect of feedback proved to be task-specific; that is, feedback facilitated perceptual learning in discrimination training and apparently emphasized the importance of stimulus discrimination throughout the experiment, whereas the effect of feedback on attentional processes and rule learning during non-discriminative control training did not transfer to post-training generalization. Admittedly, most of the other processes involved are difficult to disentangle on the basis of the present work.

It can be assumed that the same applies to the effects of discrimination training of previous studies (Ginat-Frolich, Gendler, et al., 2019; Ginat-Frolich et al., 2017; Lommen et al., 2017). They, too, most likely did not result exclusively from improved perceptual discrimination performance, although Ginat-Frolich et al. (2017) were even able to validate an increase in discrimination performance by an independent test. But first, perceptual learning per se involves several processes (Dosher & Lu, 2017), and second, the control tasks developed by Ginat-Frolich et al. (2017) or Lommen et al. (2017) differ substantially from their discrimination trainings in terms of trial structure

and time course. It is therefore questionable whether these trainings equally involve several (higher-level) processes that might facilitate perceptual discrimination in the subsequent test, e.g., working memory and attentional control.

In the studies conducted as part of this dissertation, this question can be answered with yes, which can be highlighted as strength of this work because it allowed the systematic investigation of single training aspects. So, despite the fact that some of the compared training conditions did not differ as much as expected, one can see the influence of certain aspects of the training on the reduction of fear generalization, namely that fear relevance plays a rather minor role in healthy, but could enhance training effects in individuals at risk for an anxiety disorder, and that feedback has a task-specific effect (i.e., only when given to perceptual performances).

Overall, it would be worthwhile to also systematically investigate further mechanisms that contribute to the effect of a training on fear reduction. After all, generalization can also be induced by different mechanisms. Besides similarity-based generalization (for a review see Dymond et al., 2015), there is also rule-based (Boddez et al., 2017; Livesey & McLaren, 2009; Wong & Lovibond, 2017) and category-based generalization (for a review see Dunsmoor & Murphy, 2015). Thereby, perceptual processing is not necessarily the determinant process (Glenn, Fox, Pine, Peters, & Michalska, 2020). In their fMRI study in a child sample, Glenn et al. (2020) found greater differentiation of neural patterns between CS+ and GS in affective brain areas (e.g., vmPFC, AIC, dmPFC, amygdala) compared to perceptual brain areas (e.g., inferior temporal cortex ITC, visual areas V1 and V4). Moreover, compared to low anxiety, high anxiety was associated with lower differentiation of neural patterns between stimuli only in affective brain areas (i.e., the vmPFC), whereas anxiety was not associated with altered brain

activation in perceptual areas. In this respect, it is even desirable for trainings to specifically target other processes as well.

4.3 Fear generalization and other risk factors for anxiety disorders – focusing on the individuum

Since this dissertation aimed to reduce fear generalization after its occurrence, equal to a therapeutic intervention, it was only natural to determine generalization with a focus on the individuum. After all, an intervention is not necessarily intended to help a group, but primarily to help the individual. Accordingly, I decided to aggregate every individual's response to stimuli of generalization phase by means of the Generalization Index (GI; Lenaert et al., 2016). Thus, every individual has his or her own GI value representing his or her individual generalization pattern. The index is also so well suited because it reflects exactly what is being trained, namely the discriminability and responsiveness to the GSs relative to CS+. A further advantage of GI is the combination of different components, with which a generalization gradient can be described. These include the steepness of the gradient, the mean response level as well as the range between minimum and maximum response (regularly the CS difference; Stegmann et al., 2019). However, the use of the GI in the studies presented here has revealed that the index appears to be very sensitive to one component, namely the response to CS+ (as part of the CS difference), which must be acknowledged as weakness of GI. This was particularly noticeable in Study 2, in which the groups with feedback showed more generalization (i.e., higher GI scores) pre-training, which was likely related to poorer CS differentiation at the end of fear acquisition compared to the groups without feedback. Furthermore, GI values are in a limited range of values, since they can only go towards zero, but cannot be zero or below, favoring a floor effect (Jennings & Cribbie, 2016). This also means that high values have a greater potential to decrease,

whereas low values decrease only slightly when approaching the minimum, as stated by the principle of initial values (Jennings & Cribbie, 2016). This could lead to an overestimation of the assumption that greater training needs are associated with greater success. Fortunately, the robust discrimination-training effects for US-expectancy ratings were not found to be driven substantially by observed differences pre-training (as also discussed in section 3.4). Thus, I can conclude that weaknesses of the GI arising from limited range of values and sensitivity to CS+ did not negatively affect my results. Nevertheless, it remains to be seen that the GI values like the generalization patterns of individual persons can differ considerably from each other, which leads to a large individual variance and even to outliers. Some of these were excluded from Study 2 based on statistical as well as theoretical reasons. Nonetheless, a few individuals may have favored a small effect, such as the differential benefit of different training conditions in at-risk individuals with low CS differentiation at the affective level, which should therefore be interpreted with caution. The variability of the GI is also large within an individual. This is not surprising, since responses to stimuli are aggregated per individual and the index is therefore very sensitive to response changes to a single stimulus (especially CS+). Therefore, generalization indices based on generalization patterns of individuals have lower retest-reliability than generalization patterns determined at the group level. This is undesirable for considering fear generalization as a risk factor. Nevertheless, the focus on the individual process in fear generalization is certainly desirable when it comes to transferring it to the clinic. Not for nothing, the paper by Fisher et al. (2018) titles that the “lack of group-to-individual generalizability is a threat to human subjects research”. Also, studies in the field of fear generalization point out that the mean generalization gradient of a group (e.g., the whole sample) may not be informative or even misleading, as subgroups of participants may have qualitatively different gradients (Wong & Lovibond, 2017; Zaman et al., 2021).

For this reason, of course, also influence of further risk factors of anxiety disorder (i.e. trait anxiety or deficient threat-safety differentiation) were considered as continuous variables, emphasizing the view of risk factors as an individual trait, in line with conceptualization of psychopathology as a continuum (Cuthbert, 2014; Insel et al., 2010). Indeed, in contrast to preceding studies that mostly compared mean generalization gradients of groups with differing degree of anxiety (Baumann et al., 2017; Sep et al., 2019), this approach revealed different results: anxiety increased response level, but did not influence the gradient's shape. This might be because effects are easier to find in extreme group designs, risking effect overestimation (Fisher et al., 2018; Preacher et al., 2005). Other explanations for the lack of association between generalization and anxiety could lie in the reliability of the data. For example, the high sensitivity of individual GI to extraordinary response to single stimuli, which I critically highlighted earlier in this chapter, could cause the GI not to represent a person's true value for generalization. Alternatively, the variance in anxiety might have been too low as a result of the strict inclusion criteria that did not allow individuals with current or previous clinically relevant mental health symptoms to participate. Therefore, the different results need not be understood as a contradiction. After all, increased level to all stimuli can be interpreted as generalization to even safety stimuli (Lissek et al., 2009). Possibly, both the general response level and the shape of the generalization gradient are influenced by trait anxiety, possibly depending on its severity.

This dissertation also revealed individual differences in fear generalization and its reduction depending on individual threat-safety differentiation. Specifically, poor CS differentiation appears to be a risk factor for greater generalization despite training and two generalization tests, against which relevant stimuli or feedback might be able to compensate. To my knowledge, there has been no systematic study of CSdiff at the

cognitive level (indicating contingency awareness) at the end of acquisition on subsequent fear generalization. The same is true for CSdiff at the affective level.

In conclusion, considering the individual level is important because it reveals aspects that are easily missed in group-level analyses, where interactions of risk factors with the generalization pattern can be found more easily. Considering a lack of group-to-individual generalizability, my analyses suggest that for diagnostic purposes, where the focus is on the individuum, the general response level as well as threat vs. safety differentiation may be more discriminative than the steepness of the generalization gradient (Imholze et al., in press; Stegmann et al., 2019). Moreover, taking these individual differences into account could support the choice of a most promising intervention.

4.4 Limitations and outlook

In essence, the here presented dissertation has two main limitations. Both have already been addressed several times. Firstly, participants of the collected samples are very healthy. This is also true for Study 2, even though it investigated the influence of an increased risk (e.g., trait anxiety) for an anxiety disorder on generalization and its reduction. However, the sample's individual variation in risk is not sufficient to directly infer clinical benefit of training, as only individuals without current or previous clinically relevant psychological symptoms were included. Secondly, there are several processes involved in fear generalization and its reduction, which are difficult to disentangle.

In the following, I will provide an outlook on what modifications could be made in future studies based on these limitations, while also addressing some minor limitations.

Regarding the first limitation, it is known that overgeneralization of conditioned fear particularly occurs in patients suffering from an anxiety disorder (Cooper et al., 2022; Fraunfelter et al., 2022). Nevertheless, using the same generalization paradigm as me, Wurst et al. (2021) did not observe more pronounced fear generalization in depressive or anxious-depressive patients compared to healthy controls. Therefore, it could be worthwhile to modify some aspects of the here presented generalization paradigm in order to increase ambiguity of stimuli and consequently also the extent of fear generalization.

For example, ambiguity of stimuli rises by splitting the paradigm to several days. Due to memory imprecision, generalization of US expectancy to a novel face stimulus (similar to CS+) is stronger if tested one week after single-cue fear conditioning to the CS+ face stimulus compared to a generalization test immediately following conditioning phase (Leer, Sevenster, & Lommen, 2019).

Moreover, generalization pre training might be more pronounced if participants are less experienced with the stimuli. Greater experience with the stimulus dimension may result in reduced generalization, as evidenced by a steepening of the slope (Derenne, 2019). For this reason, one could think about obtaining online instead of offline ratings in this paradigm. In this way, a further presentation of the stimuli per block could be dispensed with. Online ratings would also have the advantage that training effects would be ascertainable in the first trials of the second generalization test. This would be interesting because the generalization pattern continues to change as the phase progresses (Reutter & Gamer, 2022), and therefore the training effects would be most evident at the beginning of the phase. In this context, the question can also be raised to what extent the fixed order of stimuli during rating blocks influences the generalization gradient. Ratings per se and particularly the fixed order of stimuli

enhance the experience with the stimulus dimension, as it could facilitate to spot the number of stimuli and to distinguish them (Derenne, 2019; Sjouwerman, Niehaus, Kuhn, & Lonsdorf, 2016). However, one disadvantage of online ratings could be that they might draw attention to US-contingencies (Lonsdorf et al., 2017). In addition, online ratings of both cognitive and affective evaluations are difficult to implement because the collection of three ratings strongly disrupts the paradigm. Therefore, I decided to assess intermittent ratings as several previous studies did (Haddad et al., 2012; Holt et al., 2014; Meulders, Vandebroek, Vervliet, & Vlaeyen, 2013; Roesmann et al., 2020).

Another possibility to increase uncertainty towards stimuli is to reduce their number. Therefore, the average generalization gradient of a group is flatter if only one of several GSs was presented to each participant (between-subject design), while the average generalization gradient is steeper if all GSs were presented to each participant in a within-subjects design (Vervliet, Iberico, Vervoort, & Baeyens, 2011).

Conversely, this also means that increasing the number of stimuli during discrimination training might help reducing fear in the subsequent test. With this, I want to come to further possible modifications of the discrimination training that may increase its fear-reducing effects.

Apart from the number of stimuli, also the stimulus range on the dimensional continuum influences the generalization gradient (Derenne, 2019). Generalization becomes broader as the stimulus range increases (Hansen, Tomie, Thomas, & Thomas, 1974). The use of a wide stimulus range in discrimination training could counteract this phenomenon and possibly reduce generalization particularly effectively. It would even be conceivable to train the discrimination of stimuli that go beyond the stimulus range of the paradigm, e.g., a morph stimulus consisting of 80% CS+ and 20% of a new face,

i.e., without parts of the CS-. Supportively, fear extinction was found to be maximized when performed with a peak generalization stimulus (Struyf, Hermans, & Vervliet, 2018), which may also be true for the reduction of fear generalization. The same advantage would also emerge with training of irrelevant stimuli. Extending the dimension beyond the reference stimulus promotes more precise recognition of the same and promotes rule formation that any discrimination from the reference stimulus is beneficial.

Furthermore, a modification of trials would be conceivable. In the studies presented here, the discrimination training trials had a fixed sequence of stimulus presentations (always starting with CS+ or Line1). This might have led to an overestimation of perceptual learning (Garcia-Perez & Alcala-Quintana, 2020). Accordingly, variation in the order or discrimination of other stimulus pairs not including CS+ could increase training effects. Nevertheless, there are good reasons for my design of the training. Anxiety patients do not show sufficiently differentiated responding toward a threat stimulus (CS+) and safe similar stimuli (for meta-analyses see Cooper et al., 2022; Fraunfelter et al., 2022). Among other things, this is a result of an inability to perceptually discriminate the stimuli (Holt et al., 2014; for a review see Zaman et al., 2021). Therefore, my discrimination training was designed to specifically train the discrimination from CS+.

At last, an increase of the trial number could also enhance the fear-reducing effect of discrimination training. In both studies 1 and 2, effects of discrimination training were observed primarily for US-expectancy ratings. Apparently, new learning is more easily and earlier recognized at the cognitive level than at the affective level (W. Hofmann et al., 2010; Vansteenwegen et al., 2006). Accordingly, future studies should

investigate whether and at what level extensive discrimination training can also reduce generalization as reflected in affective responses.

The second limitation, that the mechanisms contributing to the reduction of fear generalization are difficult to disentangle, has already been discussed in detail earlier in this general discussion (see section 4.2). In providing an outlook on how this limitation might be reduced in the future, I would first propose to add a waiting group, i.e., a group that undergoes both generalization tests without intervening training. This would allow to control for processes that were equally involved in the discrimination training and the non-discriminative control training as a consequence of the decision to ensure high comparability of all training conditions, e.g., in terms of experimental structure and timing. In addition, with a waiting group one could determine to which extent safety and discrimination learning during the second generalization test further reduces fear generalization. In the course of this test, all stimuli are presented to the participants several times, so that they gain further experience with the whole dimension of interest, likely resulting in less fear generalization (Derenne, 2019; Vervliet et al., 2011).

Some processes could be studied in more detail using standardized tests. For example, a test of perceptual threshold before and after the experiment (as performed by Resnik et al., 2011; and Shalev et al., 2018) would provide information on whether and how much participants' perceptual performance improves depending on training condition.

In examining the role of executive function (or rather attentional control; Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011) in fear generalization, Niederstrasser, Meulders, Meulders, Struyf, and Vlaeyen (2017) used the stop-signal task (Verbruggen & Logan, 2008) as a measure of inhibitory capacity. Their results suggest that low inhibitory capacity is associated with slower extinction of generalized

fear. Similarly, in the present paradigm, one could examine the relationship between attentional control and fear reduction after training or the relationship between attentional control and training performance itself. However, test scores often fail to reflect individual differences in attentional abilities because they are a measure of performance effectiveness, which is less impaired by anxiety than processing efficiency (Derakshan & Eysenck, 2009). It can be concluded that the influence of attentional processes on fear generalization and its reduction can also be particularly well determined by measures of processing efficiency. One such measure is changes in brain activation measured by functional magnetic resonance imaging (fMRI). Stronger activation in the dorsolateral prefrontal cortex (dlPFC) and the ventrolateral prefrontal cortex (vlPFC) during the fear generalization paradigm (including the training) or during a standardized attention test would indicate an efficient involvement of attentional control (Derakshan & Eysenck, 2009; Ettinger et al., 2008). Stronger activation in the visual cortex during the fear generalization paradigm (including the training) or a test of perceptual threshold on the other hand would indicate an efficient involvement of early perceptual processes (in analogy with Laufer et al., 2016).

Although it would be worthwhile to systematically investigate further processes involved in fear generalization and its reduction, some of them have already been successfully investigated in this dissertation, e.g., the effects of fear relevance and feedback during (discrimination) training, as well as the vulnerability factors trait anxiety and the ability to distinguish fear from safety (for a summary of main results see, e.g., section 4).

To further improve our understanding of the role of fear generalization in the development and maintenance of anxiety disorders in the future, studies examining the predictive value of fear generalization for changes in anxiety at later time points are

needed (Scheveneels et al., 2021). To date, there is only one published study demonstrating that broader generalization to stimuli that resemble safety rather than threat predicts higher anxiety 6 months later (Lenaert et al., 2014). One study I collaborated on suggests that individual differences in basic experimental fear measures such as mean response level and CS differentiation better predict changes in negative affect following the COVID-19 pandemic outbreak, whereas individual generalization indices have no predictive value (Imholze et al., in press). In line, this dissertation showed that the inability to clearly discriminate conditioned threat and safety stimuli is a predictor of greater fear generalization despite discrimination training. However, it is possible that discrimination training with fear-relevant stimuli and feedback could help compensate for this weakness. Investigating the predictive value of experimental fear measures for the development of anxiety over time may therefore also help to better assess which training interventions might be useful for reducing maladaptive fear and anxiety.

4.5 Clinical implications

Anxiety disorders contribute greatly to the mental health burden in the United States and Europe (Bandelow & Michaelis, 2015; Wittchen et al., 2011). Understanding the key processes underlying the development of anxiety disorders will help identify at-risk individuals and optimize treatment and therefore is of great clinical relevance (Pittig, Treanor, LeBeau, & Craske, 2018). The review by Pittig et al. (2018) showed that aversive associative learning provides explanatory pathways through which anxiety and fear emerge, spread, and persist, but that additional research is needed to examine changes in aversive associative learning processes before to after treatment.

My dissertation aims to contribute to this effort by investigating for the first time whether perceptual discrimination, as one of the mechanisms contributing to fear

generalization (Holt et al., 2014; Laufer et al., 2016; Struyf et al., 2017), can be specifically trained to reduce already developed generalization in healthy participants. In doing so, I also examined other risk factors of anxiety disorders, namely trait anxiety and individual differences in fear learning (Lonsdorf & Merz, 2017; Mineka & Zinbarg, 2006), which accompany strong fear generalization (Sep et al., 2019), as affected individuals have a particular need for fear-reducing interventions but are unlikely to respond equally well to all training conditions.

What clinical implications can be drawn from my results? First, it was shown that fear generalization (pre training) is more pronounced the greater the expression of additional risk factors, at least for the risk factor deficient fear vs. safety learning. Both poor differentiation of threat and safety (for meta-analyses see Duits et al., 2015; Lissek et al., 2005) and exaggerated fear generalization (for meta-analyses see Cooper et al., 2022; Fraunfelter et al., 2022) were found to robustly occur in anxiety patients. Accordingly, training aimed at reducing fear generalization might be a promising treatment approach. However, further research on the involvement of fear generalization in the development of pathological anxiety is needed to better infer the potential benefits of trainings counteracting generalization (also see section 4.4). The prerequisite for the clinical use of such training would be met, as my dissertation has shown that the reduction of fear generalization is possible before to after discrimination training, to varying degrees depending on the training conditions and individual risk profile.

A limiting aspect is that the presented training effects seem to be largely limited to raising awareness of which stimuli predict a threat and which are safe. Anxiety patients do indeed have difficulties with this (Lissek et al., 2005; Lissek et al., 2009), but also experience fear and exhibit fear behaviors despite such awareness (Pittig, Boschet, Gluck, & Schneider, 2021). However, fear reduction at the affective level

could not be achieved with the presented trainings. Therefore, modifications are desirable to achieve fear reduction on the affective rating level as well (as already mentioned in section 4.4). Nevertheless, the present training effects on fear reduction at the cognitive rating level are an important first step, as they support achieving awareness of maladaptive avoidance behaviors and exaggerated feelings of anxiety, thus increasing motivation for therapy and therapy engagement.

In my studies, high levels of anxiety were not related to particularly strong fear generalization in a sense of a flat generalization gradient (also see Wurst et al., 2021, and section 6.4) and therefore affected individuals did not particularly benefit by discrimination training. Consequently, in anxiety patients, no general therapeutic benefit of discrimination training can be assumed. Nevertheless, there were individuals (i.e., individuals with difficulties in distinguishing threatening from safe stimuli) who reacted particularly positively to the training, especially if fear-relevant or including feedback. Given the consistently found evidence that a variety of processes are involved in the generation and reduction of excessive fear, and that individuals respond differently to these processes depending on personality and learning experience, one comes to the conclusion that individualized training would provide the greatest benefit.

As described above, training with relevant stimuli might be more beneficial for individuals who show difficulty in discriminating between threat and safety stimuli. Alternatively, consideration could be given to tailoring training to individual differences in perceptual threshold or attentional control (e.g., working memory). For example, perceptual threshold might be particularly effectively lowered if training stimuli are adapted to individual just noticeable difference (JND, for example in Holt et al., 2014)). Individuals with weaknesses in discriminating similar stimuli might also benefit from training in which the stimuli to be discriminated are presented simultaneously. On the

other hand, sequential presentation of stimuli particularly addresses working memory processes.

This way perceptual discrimination training may be a useful treatment add-on for a subgroup of patients having problems with perceptual discrimination of stimuli or with differential responding to similar stimuli. Because feared stimuli in real life are usually complex and have multiple dimensions, fear generalization also occurs on the basis of conceptual knowledge and inferential reasoning (Dunsmoor & Murphy, 2015). Therefore, a difficulty in the clinical application of training could be to account for the complexity of fear stimuli, as only the perceptual similarity dimension and single stimulus features can be targeted for training. Also, depending on the patient, the implementation of fear-relevant training is likely to be challenging.

It is also still largely unclear whether discrimination training can achieve reduction of generalization only or also anxiety symptoms in general. The results of Ginat-Frolich, Klein, et al. (2019) are optimistic, as spider phobics who underwent discrimination training compared to a non-discriminative control task showed lower avoidance of spiders in the Behavioral Avoidance Test (BAT).

Regardless of how directly applicable the discrimination training presented here is in a clinical context, the results of my dissertation make clear that individual differences in different risk factors are associated with different symptom expression (in my case, generalization), and thus individual training or treatment with individual duration is required. My analyses suggest that for diagnostic purposes, the general response level and the distinction between threat and safety may be more informative than the steepness of the generalization gradient, because the latter seems to be less reliable. In general, accounting for individual differences provides the clinic/insurance

carrier with the opportunity to select the most promising treatment and to estimate the cost, i.e., the cost calculation.

4.6 Conclusions

My dissertation investigated for the first time whether perceptual discrimination, as one of the mechanisms contributing to fear generalization (Holt et al., 2014; Laufer et al., 2016; Struyf et al., 2017), can be specifically trained to reduce already developed generalization in healthy participants. Because patients typically seek therapy at this time, this approach mimics a therapeutic intervention.

Since training was used equal to a therapeutic intervention that is intended to help individuals, generalization was determined with a focus on the individuum and individual differences in anxiety disorder risk were accounted for at a continuous level, consistent with the conceptualization of psychopathology as a continuum (Cuthbert, 2014; Insel et al., 2010).

Indeed, results successfully demonstrate a reduction in fear generalization of US expectancy, i.e., the cognitive verbal level of emotional response, before to after discrimination training. Specifically, reinforcing feedback during discrimination training consistently reduced generalization, presumably via motivational mechanisms that specifically increase engagement in perceptual tasks. On the other hand, the fear relevance of discrimination training did not prove to be particularly fear-reducing in healthy participants but could potentially enhance training effects in individuals at risk of anxiety disorder (i.e., with deficits in discriminating threat from safety at the affective level reflected in arousal ratings). These are first indications of the success of discrimination training depending on the severity of risk, which is consistent with the idea of more individualized therapy (S. G. Hofmann & Hayes, 2019).

It should be noted, however, that in addition to perceptual discrimination, a variety of processes are involved in the development and reduction of excessive fear. Given the additional evidence that individuals respond differently to these processes depending on trait anxiety and fear vs. safety learning, and that generalization differs at the cognitive vs. affective level, it is concluded that individualized training would provide the greatest benefit.

From this, it can be inferred for the clinical context that the discrimination training presented here, although not directly applicable to anxiety patients in general, might have positive effects as part of an individualized therapy for patients with difficulties in distinguishing similar stimuli.

5 References

Ahrens, L. M., Pauli, P., Reif, A., Muhlberger, A., Langs, G., Aalderink, T., & Wieser, M. J. (2016). Fear conditioning and stimulus generalization in patients with social anxiety disorder. *J Anxiety Disord, 44*, 36-46. doi:10.1016/j.janxdis.2016.10.003

Amitay, S., Zhang, Y. X., Jones, P. R., & Moore, D. R. (2014). Perceptual learning: top to bottom. *Vision Res, 99*, 69-77. doi:10.1016/j.visres.2013.11.006

Andreatta, M., Muhlberger, A., Yarali, A., Gerber, B., & Pauli, P. (2010). A rift between implicit and explicit conditioned valence in human pain relief learning. *Proc Biol Sci, 277*(1692), 2411-2416. doi:10.1098/rspb.2010.0103

Andreatta, M., & Pauli, P. (2017). Learning mechanisms underlying threat absence and threat relief: Influences of trait anxiety. *Neurobiol Learn Mem, 145*, 105-113. doi:10.1016/j.nlm.2017.09.005

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, DC: American Psychiatric Publishing.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu Rev Psychol, 63*(1), 1-29. doi:10.1146/annurev-psych-120710-100422

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in clinical neuroscience, 17*(3), 327-335. doi:10.31887/DCNS.2015.17.3/bbandelow

Banks, S. J., Bellerose, J., Douglas, D., & Jones-Gotman, M. (2012). Bilateral skin conductance responses to emotional faces. *Appl Psychophysiol Biofeedback, 37*(3), 145-152. doi:10.1007/s10484-011-9177-7

Barrett, L. F., & Russell, J. A. (2016). The Structure of Current Affect. *Current Directions in Psychological Science, 8*(1), 10-14. doi:10.1111/1467-8721.00003

Baumann, C., Schiele, M. A., Herrmann, M. J., Lonsdorf, T. B., Zwanzger, P., Domschke, K., . . . Pauli, P. (2017). Effects of an Anxiety-Specific Psychometric Factor on Fear Conditioning and Fear Generalization. *Zeitschrift Fur Psychologie-Journal of Psychology, 225*(3), 200-213. doi:10.1027/2151-2604/a000304

Beckers, T., Krypotos, A. M., Boddez, Y., Effting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biol Psychol, 92*(1), 90-96. doi:10.1016/j.biopsycho.2011.12.015

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013). Rating data are underrated: validity of US expectancy in human fear conditioning. *J Behav Ther Exp Psychiatry, 44*(2), 201-206. doi:10.1016/j.jbtep.2012.08.003

Boddez, Y., Bennett, M. P., van Esch, S., & Beckers, T. (2017). Bending rules: the shape of the perceptual generalisation gradient is sensitive to inference rules. *Cogn Emot, 31*(7), 1444-1452. doi:10.1080/02699931.2016.1230541

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., roth, W. T., Dawson, M. E., . . . Society for Psychophysiological Research Ad Hoc Committee on Electrodermal, M. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034. doi:10.1111/j.1469-8986.2012.01384.x

Bradley, M. M., & Lang, P. J. (1999). International affective digitized sounds (IADS): Stimuli, intruction manual and affective ratings. In. Gainesville, FL.

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. In R. Lane, D. & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 242-270). New York: Oxford University Press.

Brown, D. R. (1953). Stimulus-similarity and the anchoring of subjective scales. *Am J Psychol, 66*(2), 199-214. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13040526

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafo, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci, 14*(5), 365-376. doi:10.1038/nrn3475

Case, L. D., & Ambrosius, W. T. (2007). Power and sample size. *Methods Mol Biol, 404*, 377-408. doi:10.1007/978-1-59745-530-5_19

Chambless, D. L., Caputo, G. C., Bright, P., & Gallagher, R. (1984). Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire. *J Consult Clin Psychol, 52*(6), 1090-1097. doi:10.1037//0022-006x.52.6.1090

Chelazzi, L., Perlato, A., Santandrea, E., & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Res, 85*, 58-72. doi:10.1016/j.visres.2012.12.005

Constantinou, E., Purves, K. L., McGregor, T., Lester, K. J., Barry, T. J., Treanor, M., . . . Eley, T. C. (2021). Measuring fear: Association among different measures of fear learning. *J Behav Ther Exp Psychiatry, 70*, 101618. doi:10.1016/j.jbtep.2020.101618

Cooper, S. E., van Dis, E. A. M., Hagenaars, M. A., Krypotos, A. M., Nemeroff, C. B., Lissek, S., . . . Dunsmoor, J. E. (2022). A meta-analysis of conditioned fear generalization in anxiety-related disorders. *Neuropsychopharmacology, 47*(9), 1652-1661. doi:10.1038/s41386-022-01332-2

Craig, A. D., & Craig, A. (2009). How do you feel--now? The anterior insula and human awareness. *Nature reviews neuroscience, 10*(1).

Cuthbert, B. N. (2014). The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry, 13*(1), 28-35. doi:10.1002/wps.20087

Davis, M., Walker, D. L., Miles, L., & Grillon, C. (2010). Phasic vs sustained fear in rats and humans: role of the extended amygdala in fear vs anxiety. *Neuropsychopharmacology, 35*(1), 105-135. doi:10.1038/npp.2009.109

Derakshan, N., & Eysenck, M. W. (2009). Anxiety, Processing Efficiency, and Cognitive Performance New Developments from Attentional Control Theory. *European Psychologist, 14*(2), 168-176. doi:10.1027/1016-9040.14.2.168

Derenne, A. (2019). Gradient Shift is Sensitive to the Width of the Generalization Test Range. *Psychological Record, 69*(3), 401-414. doi:10.1007/s40732-019-00342-9

Dosher, B., & Lu, Z. L. (2017). Visual Perceptual Learning and Models. *Annu Rev Vis Sci, 3*, 343-363. doi:10.1146/annurev-vision-102016-061249

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. *Depress Anxiety, 32*(4), 239-253. doi:10.1002/da.22353

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus Typicality Determines How Broadly Fear Is Generalized. *Psychological science, 25*(9), 1816-1821. doi:10.1177/0956797614535401

Dunsmoor, J. E., & Murphy, G. L. (2015). Categories, concepts, and conditioning: how humans generalize fear. *Trends Cogn Sci, 19*(2), 73-77. doi:10.1016/j.tics.2014.12.003

Dunsmoor, J. E., & Paz, R. (2015). Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. *Biol Psychiatry, 78*(5), 336-343. doi:10.1016/j.biopsych.2015.04.010

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear Generalization in Humans: Systematic Review and Implications for Anxiety Disorder Research. *Behav Ther, 46*(5), 561-582. doi:10.1016/j.beth.2014.10.001

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J Pers Soc Psychol, 17*(2), 124-129. doi:10.1037/h0030377

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. Englewood Cliffs, NJ: Prentice Hall.

Ettinger, U., Ffytche, D. H., Kumari, V., Kathmann, N., Reuter, B., Zelaya, F., & Williams, S. C. R. (2008). Decomposing the neural correlates of antisaccade eye movements using event-related fMRI. *Cerebral Cortex, 18*(5), 1148-1159. doi:10.1093/cercor/bhm147

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences, 50*(7), 955-960. doi:10.1016/j.paid.2010.08.019

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion, 7*(2), 336-353. doi:10.1037/1528-3542.7.2.336

Fishbach, A., Eyal, T., & Finkelstein, S. R. (2010). How Positive and Negative Feedback Motivate Goal Pursuit. *Social and Personality Psychology Compass, 4*(8), 517-530. doi:10.1111/j.1751-9004.2010.00285.x

Fisher, A. J., Granger, D. A., & Newman, M. G. (2010). Sympathetic arousal moderates self-reported physiological arousal symptoms at baseline and physiological flexibility in response to a stressor in generalized anxiety disorder. *Biological Psychology, 83*(3), 191-200. doi:10.1016/j.biopsycho.2009.12.007

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences, 115*(27), E6106-E6115.

Fraunfelter, L., Gerdes, A. B. M., & Alpers, G. W. (2022). Fear one, fear them all: A systematic review and meta-analysis of fear generalization in pathological anxiety. *Neurosci Biobehav Rev, 139*, 104707. doi:10.1016/j.neubiorev.2022.104707

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Avila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry, 21*(4), 500-508. doi:10.1038/mp.2015.88

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: object specificity and size invariance. *Vision Res, 40*(5), 473-484. doi:10.1016/s0042-6989(99)00134-0

Fusar-Poli, P., Landi, P., & O'Connor, C. (2009). Neurophysiological response to emotional faces with increasing intensity of fear: a skin conductance response study. *J Clin Neurosci, 16*(7), 981-982. doi:10.1016/j.jocn.2008.09.022

Fydrich, T. (1999). SPAI–Soziale Phobie und Angst Inventar. Deutschsprachige Adaption des “Social Phobia and Anxiety Inventory” von Turner und Beidel. *Unpublished manuscript, Psychological Institute, University of Heidelberg, Germany*.

Garcia-Perez, M. A., & Alcala-Quintana, R. (2020). Order effects in two-alternative forced-choice tasks invalidate adaptive threshold

estimates. *Behav Res Methods, 52*(5), 2168-2187. doi:10.3758/s13428-020-01384-6

Gazendam, F. J., Kamphuis, J. H., & Kindt, M. (2013). Deficient safety learning characterizes high trait anxious individuals. *Biol Psychol, 92*(2), 342-352. doi:10.1016/j.biopsycho.2012.11.006

Ginat-Frolich, R., Gendler, T., Marzan, D., Tsuk, Y., & Shechner, T. (2019). Reducing fear overgeneralization in children using a novel perceptual discrimination task. *Behav Res Ther, 116*, 131-139. doi:10.1016/j.brat.2019.03.008

Ginat-Frolich, R., Klein, Z., Aderka, I. M., & Shechner, T. (2019). Reducing avoidance in adults with high spider fear using perceptual discrimination training. *Depress Anxiety, 36*(9), 859-865. doi:10.1002/da.22930

Ginat-Frolich, R., Klein, Z., Katz, O., & Shechner, T. (2017). A novel perceptual discrimination training task: Reducing fear overgeneralization in the context of fear learning. *Behav Res Ther, 93*, 29-37. doi:10.1016/j.brat.2017.03.010

Glenn, D. E., Fox, N. A., Pine, D. S., Peters, M. A. K., & Michalska, K. J. (2020). Divergence in cortical representations of threat generalization in affective versus perceptual circuitry in childhood: Relations with anxiety. *Neuropsychologia, 142*, 107416. doi:10.1016/j.neuropsychologia.2020.107416

Goodman, J. S., Wood, R. E., & Chen, Z. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behavior and Human Decision Processes, 115*(2), 253-267. doi:10.1016/j.obhdp.2011.01.001

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *J Exp Psychol, 51*(1), 79-88. doi:10.1037/h0046219

Haddad, A. D., Pritchett, D., Lissek, S., & Lau, J. Y. F. (2012). Trait Anxiety and Fear Responses to Safety Cues: Stimulus Generalization or Sensitization? *Journal of Psychopathology and Behavioral Assessment, 34*(3), 323-331. doi:10.1007/s10862-012-9284-7

Haddad, A. D., Xu, M., Raeder, S., & Lau, J. Y. (2013). Measuring the role of conditioning and stimulus generalisation in common fears and worries. *Cogn Emot, 27*(5), 914-922. doi:10.1080/02699931.2012.747428

Hamm, A. O. (2020). Fear, anxiety, and their disorders from the perspective of psychophysiology. *Psychophysiology, 57*(2), e13474. doi:10.1111/psyp.13474

Hamm, A. O., & Vaitl, D. (1996). Affective learning: awareness and aversion. *Psychophysiology, 33*(6), 698-710. doi:10.1111/j.1469-8986.1996.tb02366.x

Hansen, G., Tomie, A., Thomas, D. R., & Thomas, D. H. (1974). Effect of test stimulus range on stimulus generalization in human subjects. *J Exp Psychol, 102*(4), 634-639. doi:10.1037/h0036097

Harrison, B. J., Fullana, M. A., Via, E., Soriano-Mas, C., Vervliet, B., Martinez-Zalacain, I., . . . Cardoner, N. (2017). Human ventromedial prefrontal cortex and the positive affective processing of safety signals. *NeuroImage, 152*, 12-18. doi:10.1016/j.neuroimage.2017.02.080

Hautzinger, M., Keller, F., & Kühner, C. (2006). *BDI-II. Beck Depressions Inventar Revision - Manual*. Frankfurt: Harcourt Test Services.

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy, 40*(3), 217-234. doi:10.1016/S0005-7967(01)00006-7

Herzog, K., Andreatta, M., Schneider, K., Schiele, M. A., Domschke, K., Romanos, M., . . . Pauli, P. (2021). Reducing Generalization of Conditioned Fear: Beneficial Impact of Fear Relevance and Feedback in Discrimination Training. *Front Psychol, 12*(1648), 665711. doi:10.3389/fpsyg.2021.665711

Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision Res, 39*(25), 4232-4243. doi:10.1016/s0042-6989(99)00138-8

Hofmann, S. G., & Hayes, S. C. (2019). The Future of Intervention Science: Process-Based Therapy. *Clin Psychol Sci, 7*(1), 37-50. doi:10.1177/2167702618772296

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychol Bull, 136*(3), 390-421. doi:10.1037/a0018916

Holt, D. J., Boeke, E. A., Wolthusen, R. P., Nasr, S., Milad, M. R., & Tootell, R. B. (2014). A parametric study of fear generalization to faces and non-face objects: relationship to discrimination thresholds. *Front Hum Neurosci, 8*(624), 624. doi:10.3389/fnhum.2014.00624

Imholze, C., Hutterer, K., Gall, D., Dannlowski, U., Domschke, K., Leehr, E. J., . . . Gamer, M. (in press). Prediction of changes in negative affect during the COVID-19 pandemic by experimental fear conditioning and generalization measures - a longitudinal study. *Zeitschrift Fur Psychologie-Journal of Psychology*.

Indovina, I., Robbins, T. W., Nunez-Elizalde, A. O., Dunn, B. D., & Bishop, S. J. (2011). Fear-conditioning mechanisms associated with trait vulnerability to anxiety in humans. *Neuron, 69*(3), 563-571. doi:10.1016/j.neuron.2010.12.034

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry, 167*(7), 748-751. doi:10.1176/appi.ajp.2010.09091379

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS medicine, 2*(8), 696-701. doi:10.1371/journal.pmed.0020124

Izard, C. E. (2010). The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. *Emotion Review, 2*(4), 363-370. doi:10.1177/1754073910374661

Jennings, M., & Cribbie, R. A. (2016). Comparing pre-post change across groups: Guidelines for choosing between difference scores, ANCOVA, and residual change scores.

Kaczkurkin, A. N., Burton, P. C., Chazin, S. M., Manbeck, A. B., Espensen-Sturges, T., Cooper, S. E., . . . Lissek, S. (2017). Neural Substrates of Overgeneralized Conditioned Fear in PTSD. *Am J Psychiatry, 174*(2), 125-134. doi:10.1176/appi.ajp.2016.15121549

Kemper, C. J., Ziegler, M., & Taylor, S. (2009). Überprüfung der psychometrischen Qualität der deutschen Version des Angstsensitivitätsindex-3. *Diagnostica, 55*(4), 223-233.

Kendler, K. S., Myers, J., & Prescott, C. A. (2002). The etiology of phobias: an evaluation of the stress-diathesis model. *Arch Gen Psychiatry, 59*(3), 242-248. doi:10.1001/archpsyc.59.3.242

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5*(4), 345-379. doi:10.1007/bf00992553

Krohne, H. W., Egloff, B., Kohmann, C. W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS). *Diagnostica, 42*, 139-156.

Lang, P. J. (1968). Fear reduction and fear behavior: Problems in treating a construct. In *Research in psychotherapy.* (pp. 90-102). Washington, DC, US: American Psychological Association.

Lang, P. J., & Bradley, M. M. (2010). Emotion and the motivational brain. *Biol Psychol, 84*(3), 437-450. doi:10.1016/j.biopsycho.2009.10.007

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): affective ratings of pictures and instruction manual. University of Florida, Gainesville*. Retrieved from

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*(3), 261-273. doi:10.1111/j.1469-8986.1993.tb03352.x

Lang, P. J., & McTeague, L. M. (2009). The anxiety disorder spectrum: fear imagery, physiological reactivity, and differential diagnosis.

Anxiety Stress Coping, 22(1), 5-25. doi:10.1080/10615800802478247

Lange, I., Goossens, L., Michielse, S., Bakker, J., Lissek, S., Papalini, S., . . . Schruers, K. (2017). Behavioral pattern separation and its link to the neural mechanisms of fear generalization. *Soc Cogn Affect Neurosci, 12*(11), 1720-1729. doi:10.1093/scan/nsx104

Lau, J. Y. F., Lissek, S., Nelson, E. E., Lee, Y., Roberson-Nay, R., Poeth, K., . . . Pine, D. S. (2008). Fear conditioning in adolescents with anxiety disorders: results from a novel experimental paradigm. *J Am Acad Child Adolesc Psychiatry, 47*(1), 94-102. doi:10.1097/chi.0b01e31815a5f01

Laufer, O., Israeli, D., & Paz, R. (2016). Behavioral and Neural Mechanisms of Overgeneralization in Anxiety. *Curr Biol, 26*(6), 713-722. doi:10.1016/j.cub.2016.01.023

Laufer, O., & Paz, R. (2012). Monetary loss alters perceptual thresholds and compromises future decisions via amygdala and prefrontal networks. *J Neurosci, 32*(18), 6304-6311. doi:10.1523/JNEUROSCI.6281-11.2012

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *Das State-Trait Angstinventar.* Weinheim: Beltz Test.

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A Unified Framework to Quantify the Credibility of Scientific Findings. *Advances in Methods and Practices in Psychological Science, 1*(3), 389-402. doi:10.1177/2515245918787189

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu Rev Neurosci, 23*(1), 155-184. doi:10.1146/annurev.neuro.23.1.155

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

Leer, A., Sevenster, D., & Lommen, M. J. J. (2019). Generalisation of threat expectancy increases with time. *Cogn Emot, 33*(5), 1067-1075. doi:10.1080/02699931.2018.1526167

Leknes, S., Lee, M., Berna, C., Andersson, J., & Tracey, I. (2011). Relief as a Reward: Hedonic and Neural Responses to Safety from Pain. *PloS one, 6*(4), e17870. doi:10.1371/journal.pone.0017870

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. *J Anxiety Disord, 28*(8), 747-753. doi:10.1016/j.janxdis.2014.09.006

Lenaert, B., van de Ven, V., Kaas, A. L., & Vlaeyen, J. W. (2016). Generalization on the Basis of Prior Experience Is Predicted by

Individual Differences in Working Memory. *Behav Ther, 47*(1), 130-140. doi:10.1016/j.beth.2015.10.001

Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans'. *The American Statistician, 34*(4), 216-221.

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition & emotion, 19*(2), 269-282. doi:10.1080/02699930441000319

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: experimental validation and clinical relevance. *Behav Res Ther, 46*(5), 678-687. doi:10.1016/j.brat.2008.02.005

Lissek, S., Bradford, D. E., Alvarez, R. P., Burton, P., Espensen-Sturges, T., Reynolds, R. C., & Grillon, C. (2014). Neural substrates of classically conditioned fear-generalization in humans: a parametric fMRI study. *Soc Cogn Affect Neurosci, 9*(8), 1134-1142. doi:10.1093/scan/nst096

Lissek, S., & Grillon, C. (2010). Overgeneralization of Conditioned Fear in the Anxiety Disorders Putative Memorial Mechanisms. *Zeitschrift Fur Psychologie-Journal of Psychology, 218*(2), 146-148. doi:10.1027/0044-3409/a000022

Lissek, S., & Grillon, C. (2012). Learning models of PTSD. In J. G. Beck & D. M. Sloan (Eds.), *The Oxford handbook of traumatic stress disorders* (pp. 175–190): Oxford University Press.

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biol Psychiatry, 75*(11), 909-915. doi:10.1016/j.biopsych.2013.07.025

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. *Behav Res Ther, 43*(11), 1391-1424. doi:10.1016/j.brat.2004.10.007

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. *Am J Psychiatry, 167*(1), 47-55. doi:10.1176/appi.ajp.2009.09030410

Lissek, S., Rabin, S. J., McDowell, D. J., Dvir, S., Bradford, D. E., Geraci, M., . . . Grillon, C. (2009). Impaired discriminative fear-conditioning resulting from elevated fear responding to learned safety cues among individuals with panic disorder. *Behav Res Ther, 47*(2), 111-118. doi:10.1016/j.brat.2008.10.017

Liu, J., Lu, Z. L., & Dosher, B. A. (2010). Augmented Hebbian reweighting: interactions between feedback and training accuracy in perceptual learning. *J Vis, 10*(10), 29. doi:10.1167/10.10.29

Livesey, E. J., & McLaren, I. P. (2009). Discrimination and generalization along a simple dimension: peak shift and rule-governed responding. *J Exp Psychol Anim Behav Process, 35*(4), 554-565. doi:10.1037/a0015524

Lohr, J. M., Olatunji, B. O., & Sawchuk, C. N. (2007). A functional analysis of danger and safety signals in anxiety disorders. *Clin Psychol Rev, 27*(1), 114-126. doi:10.1016/j.cpr.2006.07.005

Lommen, M. J. J., Duta, M., Vanbrabant, K., de Jong, R., Juechems, K., & Ehlers, A. (2017). Training discrimination diminishes maladaptive avoidance of innocuous stimuli in a fear conditioning paradigm. *PloS one, 12*(10), e0184485. doi:10.1371/journal.pone.0184485

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. *Neurosci Biobehav Rev, 77*, 247-285. doi:10.1016/j.neubiorev.2017.02.026

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans - Biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience and Biobehavioral Reviews, 80*, 703-728. doi:10.1016/j.neubiorev.2017.07.007

Lueken, U., Straube, B., Reinhardt, I., Maslowski, N. I., Wittchen, H. U., Strohle, A., . . . Kircher, T. (2014). Altered top-down and bottom-up processing of fear conditioning in panic disorder with agoraphobia. *Psychol Med, 44*(2), 381-394. doi:10.1017/S0033291713000792

Marstaller, L., Al-Jiboury, R., Kemp, A. H., & Dymond, S. (2021). Rule-based generalization of threat without similarity. *Biol Psychol, 160*, 108042. doi:10.1016/j.biopsycho.2021.108042

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist, 70*(6), 487.

Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response. *Brain Struct Funct, 214*(5-6), 535-549. doi:10.1007/s00429-010-0265-x

Mendoza, J. L., Stafford, K. L., & Stauffer, J. M. (2000). Large-sample confidence intervals for validity and reliability coefficients. *Psychol Methods, 5*(3), 356-369. doi:10.1037/1082-989x.5.3.356

Mertens, G., Bouwman, V., & Engelhard, I. M. (2021). Conceptual fear generalization gradients and their relationship with anxious traits: Results from a Registered Report. *Int J Psychophysiol, 170*, 43-50. doi:10.1016/j.ijpsycho.2021.09.007

Meulders, A., Harvie, D. S., Lorimer Moseley, G., & Vlaeyen, J. W. (2015). Generalization of Pain-Related Fear Using a Left-Right Hand Judgment Conditioning Task. *Behav Ther, 46*(5), 699-716. doi:10.1016/j.beth.2015.02.004

Meulders, A., Vandebroek, N., Vervliet, B., & Vlaeyen, J. W. (2013). Generalization gradients in cued and contextual pain-related fear: an experimental study in healthy participants. *Front Hum Neurosci, 7*(345), 345. doi:10.3389/fnhum.2013.00345

Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational neuroscience: ten years of progress. *Annu Rev Psychol, 63*, 129-151. doi:10.1146/annurev.psych.121208.131631

Mineka, S., & Oehlberg, K. (2008). The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. *Acta Psychol (Amst), 127*(3), 567-580. doi:10.1016/j.actpsy.2007.11.007

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

Navarro, D. (2015). lsr: Companion to "Learning Statistics with R"(Version 0.5). *Pridobljeno iz https://cran. r-project. org/package= lsr*.

Niederstrasser, N. G., Meulders, A., Meulders, M., Struyf, D., & Vlaeyen, J. W. (2017). Executive functions deficits impair extinction of generalization of fear of movement-related pain. *Eur J Pain, 21*(5), 886-899. doi:10.1002/ejp.991

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychol Rev, 108*(3), 483-522. doi:10.1037/0033-295x.108.3.483

Onat, S. (2018). Towards a new understanding of fear generalization and its neural origin. *PeerJ Preprints, 6*, e27311v27311.

Onat, S., & Buchel, C. (2015). The neuronal basis of fear generalization in humans. *Nat Neurosci, 18*(12), 1811-1818. doi:10.1038/nn.4166

Otto, T., & Eichenbaum, H. (1992). Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: evidence for hippocampal processing in recognition memory. *Hippocampus, 2*(3), 323-334. doi:10.1002/hipo.450020310

Pashler, H., & Wagenmakers, E. J. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspect Psychol Sci, 7*(6), 528-530. doi:10.1177/1745691612465253

Pavlov, I. P. (1927). *Conditional reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford, England: Oxford Univ. Press.

Pittig, A., Boschet, J. M., Gluck, V. M., & Schneider, K. (2021). Elevated costly avoidance in anxiety disorders: Patients show little downregulation of acquired avoidance in face of competing rewards for approach. *Depress Anxiety, 38*(3), 361-371. doi:10.1002/da.23119

Pittig, A., Treanor, M., LeBeau, R. T., & Craske, M. G. (2018). The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neurosci Biobehav Rev, 88*, 117-140. doi:10.1016/j.neubiorev.2018.03.015

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology, 17*(3), 715-734. doi:10.1017/S0954579405050340

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10*(2), 178-192. doi:10.1037/1082-989x.10.2.178

Ramirez, F., Moscarello, J. M., LeDoux, J. E., & Sears, R. M. (2015). Active avoidance requires a serial basal amygdala to nucleus accumbens shell circuit. *J Neurosci, 35*(8), 3470-3477. doi:10.1523/JNEUROSCI.1331-14.2015

Raymond, J. G., Steele, J. D., & Series, P. (2017). Modeling Trait Anxiety: From Computational Processes to Personality. *Front Psychiatry, 8*, 1. doi:10.3389/fpsyt.2017.00001

Reinhard, J., Slyschak, A., Schiele, M. A., Andreatta, M., Kneer, K., Reif, A., . . . Romanos, M. (2022). Fear conditioning and stimulus generalization in association with age in children and adolescents. *Eur Child Adolesc Psychiatry, 31*(10), 1581-1590. doi:10.1007/s00787-021-01797-4

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *J Comp Physiol Psychol, 66*(1), 1-5. doi:10.1037/h0025984

Resnik, J., Sobel, N., & Paz, R. (2011). Auditory aversive learning increases discrimination thresholds. *Nat Neurosci, 14*(6), 791-796. doi:10.1038/nn.2802

Reutter, M., & Gamer, M. (2022). Individual patterns of visual exploration predict the extent of fear generalization in humans. *Emotion*. doi:10.1037/emo0001134

Roesmann, K., Wiens, N., Winker, C., Rehbein, M. A., Wessing, I., & Junghoefer, M. (2020). Fear generalization of implicit conditioned facial features - Behavioral and magnetoencephalographic correlates. *NeuroImage, 205*, 116302. doi:10.1016/j.neuroimage.2019.116302

Rolls, E. T. (2019). The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Struct Funct, 224*(9), 3001-3018. doi:10.1007/s00429-019-01945-2

Rosebrock, L. E., Hoxha, D., Norris, C., Cacioppo, J. T., & Gollan, J. K. (2017). Skin Conductance and Subjective Arousal in Anxiety, Depression, and Comorbidity. *Journal of Psychophysiology, 31*(4), 145-157. doi:10.1027/0269-8803/a000176

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J Pers Soc Psychol, 76*(5), 805-819. doi:10.1037//0022-3514.76.5.805

Samsom, D., & Rachman, S. (1989). The effect of induced mood on fear reduction. *Br J Clin Psychol, 28*(3), 227-238. doi:10.1111/j.2044-8260.1989.tb01372.x

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nat Rev Neurosci, 11*(1), 53-60. doi:10.1038/nrn2737

Schechtman, E., Laufer, O., & Paz, R. (2010). Negative valence widens generalization of learning. *J Neurosci, 30*(31), 10460-10464. doi:10.1523/JNEUROSCI.2377-10.2010

Scheveneels, S., Boddez, Y., & Hermans, D. (2021). Predicting clinical outcomes via human fear conditioning: A narrative review. *Behav Res Ther, 142*, 103870. doi:10.1016/j.brat.2021.103870

Schiele, M. A., Reinhard, J., Reif, A., Domschke, K., Romanos, M., Deckert, J., & Pauli, P. (2016). Developmental aspects of fear: Comparing the acquisition and generalization of conditioned fear in children and adults. *Dev Psychobiol, 58*(4), 471-481. doi:10.1002/dev.21393

Schultz, D. H., & Helmstetter, F. J. (2010). Classical Conditioning of Autonomic Fear Responses Is Independent of Contingency Awareness. *Journal of Experimental Psychology-Animal Behavioral Processes, 36*(4), 495-500. doi:10.1037/a0020263

Sehlmeyer, C., Schoning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., & Konrad, C. (2009). Human fear conditioning and extinction in neuroimaging: a systematic review. *PloS one, 4*(6), e5865. doi:10.1371/journal.pone.0005865

Sep, M. S. C., Steenmeijer, A., & Kennis, M. (2019). The relation between anxious personality traits and fear generalization in healthy subjects: A systematic review and meta-analysis. *Neurosci Biobehav Rev, 107*, 320-328. doi:10.1016/j.neubiorev.2019.09.029

Shalev, L., Paz, R., & Avidan, G. (2018). Visual Aversive Learning Compromises Sensory Discrimination. *J Neurosci, 38*(11), 2766-2779. doi:10.1523/JNEUROSCI.0889-17.2017

Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of*

Experimental Psychology-Animal Behavior Processes, 24(4), 405-415. doi:Doi 10.1037/0097-7403.24.4.405

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry, 59 Suppl 20*(20), 22-33.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2015). afex: Analysis of factorial experiments. *R package version 0.13–145*.

Sjouwerman, R., Niehaus, J., Kuhn, M., & Lonsdorf, T. B. (2016). Don't startle me-Interference of startle probe presentations and intermittent ratings with fear acquisition. *Psychophysiology, 53*(12), 1889-1899. doi:10.1111/psyp.12761

Sjouwerman, R., Scharfenort, R., & Lonsdorf, T. B. (2020). Individual differences in fear learning: Specificity to trait-anxiety beyond other measures of negative affect, and mediation via amygdala activation. *bioRxiv*, 233528. doi:10.1101/233528

Stangier, U., & Heidenreich, T. (2003). Die Liebowitz Soziale Angst-Skala (LSAS). In *Colloquium Internationale Psychiatriae Scalarum*. Bern: Huber.

Stegmann, Y., Ahrens, L., Pauli, P., Keil, A., & Wieser, M. J. (2020). Social aversive generalization learning sharpens the tuning of visuocortical neurons to facial identity cues. *Elife, 9*, e55204. doi:10.7554/eLife.55204

Stegmann, Y., Schiele, M. A., Schumann, D., Lonsdorf, T. B., Zwanzger, P., Romanos, M., . . . Pauli, P. (2019). Individual differences in human fear generalization-pattern identification and implications for anxiety disorders. *Transl Psychiatry, 9*(1), 307. doi:10.1038/s41398-019-0646-8

Struyf, D., Hermans, D., & Vervliet, B. (2018). Maximizing the generalization of fear extinction: Exposures to a peak generalization stimulus. *Behav Res Ther, 111*, 1-8. doi:10.1016/j.brat.2018.09.005

Struyf, D., Zaman, J., Hermans, D., & Vervliet, B. (2017). Gradients of fear: How perception influences fear generalization. *Behav Res Ther, 93*, 116-122. doi:10.1016/j.brat.2017.04.001

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

Taschereau-Dumouchel, V., Michel, M., Lau, H., Hofmann, S. G., & LeDoux, J. E. (2022). Putting the "mental" back in "mental disorders": a perspective from research on fear and anxiety. *Mol Psychiatry, 27*(3), 1322-1330. doi:10.1038/s41380-021-01395-5

Taylor, S., Zvolensky, M. J., Cox, B. J., Deacon, B., Heimberg, R. G., Ledley, D. R., . . . Cardenas, S. J. (2007). Robust dimensions of anxiety sensitivity: development and initial validation of the Anxiety Sensitivity Index-3. *Psychol Assess, 19*(2), 176-188. doi:10.1037/1040-3590.19.2.176

Tinoco-Gonzalez, D., Fullana, M. A., Torrents-Rodas, D., Bonillo, A., Vervliet, B., Blasco, M. J., . . . Torrubia, R. (2015). Conditioned Fear Acquisition and Generalization in Generalized Anxiety Disorder. *Behav Ther, 46*(5), 627-639. doi:10.1016/j.beth.2014.12.004

Torrents-Rodas, D., Fullana, M. A., Bonillo, A., Caseras, X., Andion, O., & Torrubia, R. (2013). No effect of trait anxiety on differential fear conditioning or fear generalization. *Biol Psychol, 92*(2), 185-190. doi:10.1016/j.biopsycho.2012.10.006

Tracy, J. L., & Randles, D. (2011). Four Models of Basic Emotions: A Review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3*(4), 397-405. doi:10.1177/1754073911410747

Turner, S. M., Beidel, D. C., Dancu, C. V., & Stanley, M. A. (1989). An empirically derived inventory to measure social fears and anxiety: The Social Phobia and Anxiety Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 35-40. doi:10.1037/1040-3590.1.1.35

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *J Exp Psychol Anim Behav Process, 32*(1), 71-79. doi:10.1037/0097-7403.32.1.71

Veit, R., Flor, H., Erb, M., Hermann, C., Lotze, M., Grodd, W., & Birbaumer, N. (2002). Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Neurosci Lett, 328*(3), 233-236. doi:10.1016/s0304-3940(02)00519-0

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends Cogn Sci, 12*(11), 418-424. doi:10.1016/j.tics.2008.07.005

Vervliet, B., Iberico, C., Vervoort, E., & Baeyens, F. (2011). Generalization gradients in human predictive learning: Effects of discrimination training and within-subjects testing. *Learning and Motivation, 42*(3), 210-220. doi:10.1016/j.lmot.2011.03.004

Vogt, B. A. (2009). *Cingulate neurobiology and disease*: Oxford University Press.

Vogt, B. A., & Laureys, S. (2009). The primate posterior cingulate gyrus: connections, sensorimotor orientation, gateway to limbic processing. In B. A. Vogt (Ed.), *Cingulate neurobiology and disease* (pp. 275-308). New York: Oxford University Press.

Weil, R. S., Furl, N., Ruff, C. C., Symmonds, M., Flandin, G., Dolan, R. J., . . . Rees, G. (2010). Rewarding feedback after correct visual discriminations has both general and specific influences on visual cortex. *J Neurophysiol, 104*(3), 1746-1757. doi:10.1152/jn.00870.2009

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology, 26*(4), 557-580. doi:Doi 10.1002/(Sici)1099-0992(199607)26:4<557::Aid-Ejsp769>3.0.Co;2-4

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jonsson, B., . . . Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol, 21*(9), 655-679. doi:10.1016/j.euroneuro.2011.07.018

Wong, A. H. K., & Lovibond, P. F. (2017). Rule-based generalisation in single-cue and differential fear conditioning in humans. *Biol Psychol, 129*, 111-120. doi:10.1016/j.biopsycho.2017.08.056

Wurst, C., Schiele, M. A., Stonawski, S., Weiss, C., Nitschke, F., Hommers, L., . . . Menke, A. (2021). Impaired fear learning and extinction, but not generalization, in anxious and non-anxious depression. *J Psychiatr Res, 135*, 294-301. doi:10.1016/j.jpsychires.2021.01.034

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends Neurosci, 34*(10), 515-525. doi:10.1016/j.tins.2011.06.006

Zaman, J., Ceulemans, E., Hermans, D., & Beckers, T. (2019). Direct and indirect effects of perception on generalization gradients. *Behav Res Ther, 114*, 44-50. doi:10.1016/j.brat.2019.01.006

Zaman, J., Chalkia, A., Zenses, A. K., Bilgin, A. S., Beckers, T., Vervliet, B., & Boddez, Y. (2021). Perceptual variability: Implications for learning and generalization. *Psychon Bull Rev, 28*(1), 1-19. doi:10.3758/s13423-020-01780-1

Zbozinek, T. D., Holmes, E. A., & Craske, M. G. (2015). The effect of positive mood induction on reducing reinstatement fear: Relevance for long term outcomes of exposure therapy. *Behav Res Ther, 71*, 65-75. doi:10.1016/j.brat.2015.05.016

Zvolensky, M. J., Kotov, R., Antipova, A. V., & Schmidt, N. B. (2005). Diathesis stress model for panic-related distress: a test in a Russian epidemiological sample. *Behav Res Ther, 43*(4), 521-532. doi:10.1016/j.brat.2004.09.001

6 Appendix

6.1 Study 1A

6.1.1 State questionnaires

As state control measures, the state version of the STAI and the Positive and Negative Affective Schedule (PANAS, Krohne, Egloff, Kohmann, & Tausch, 1996) were completed at the beginning and the end of the experimental protocol (**Supplementary Table 1** and **Supplementary Table 2**).

Supplementary Table 1 *Changes in the emotional state of the participants throughout the experimental procedure independently from the task*

	STAI state	*Negative Mood*	*Positive Mood*
begin (*SD*)	32.90 (6.37)	11.50 (2.25)	31.74 (5.85)
end (*SD*)	36.51 (8.06)	13.05 (3.31)	30.08 (6.36)

Notes. STAI State Trait Anxiety Inventory, *Negative Mood* and *Positive Mood* subscales of the Positive and Negative Affect Schedule (PANAS).

The 2 (phase: beginning, end of experiment) × 2 (training: relevant_DT, irrelevant_DT) × 2 (feedback: with, without) ANOVAs show a significant main effect for phases for all three scales (**Supplementary Table 1**; state anxiety: $F(1,76) = 27.28$, $p < .001$, $\eta_p^2 = .264$; negative mood: $F(1,76) = 20.96$, $p < .001$, $\eta_p^2 = .216$; positive mood: $F(1,76) = 11.46$, $p = .001$, $\eta_p^2 = .131$) indicating that participants were more anxious ($F(1,76) = 27.28$, $p < .001$, $\eta_p^2 = .264$) and had higher negative mood ($F(1,76) = 20.96$, $p < .001$, $\eta_p^2 = .216$) and lower positive mood ($F(1,76) = 11.46$, $p = .001$, $\eta_p^2 = .131$) at the end of the experiment compared to the beginning. I also observed a significant interaction between training and feedback for the negative mood scale ($F(1,76) = 5.72$, $p = .019$, $\eta_p^2 = .070$), and a significant Phase × Training × Feedback

interaction (**Supplementary Table 2**) for state anxiety ($F(1,76) = 4.63$, $p = .035$, $\eta_p^2 = .057$) and negative mood ($F(1,76) = 7.47$, $p = .008$, $\eta_p^2 = .089$), but not for positive mood ($F(1,76) = 0.15$, $p = .704$, $\eta_p^2 = .002$). No other effects were found (all p values > .069). Post-hoc simple contrasts (Bonferroni corrected, $\alpha < .012$) for the three-way interaction showed that participants who received fear-irrelevant discrimination training with feedback had higher state anxiety ($F(1,19) = 16.83$, $p = .001$, $\eta_p^2 = .470$) and higher negative mood ($F(1,19) = 13.95$, $p = .001$, $\eta_p^2 = .423$) at the end of the experiment compared to the beginning, while all other groups showed no significant change in their anxiety levels or mood throughout the experiment (all p values > .016).

Supplementary Table 2 *Changes in the emotional state of the participants throughout the experimental procedure dependent on the task*

	relevant_DT _noFB	*relevant_DT _FB*	*irrelevant_DT _noFB*	*irrelevant_DT _FB*
STAI state				
begin (*SD*)	34.40 (6.11)	33.80 (5.83)	31.10 (7.38)	32.30 (6.00)
end (*SD*)	38.05 (6.64)	35.65 (5.62)	33.50 (8.12)	38.85 (10.48)
Negative Mood				
begin (*SD*)	11.95 (2.06)	11.35 (2.30)	11.35 (1.98)	11.35 (2.70)
end (*SD*)	13.50 (3.15)	11.65 (1.63)	12.30 (2.39)	14.75 (4.61)
Positive Mood				
begin (*SD*)	31.15 (4.60)	33.80 (5.28)	31.45 (6.07)	30.55 (7.06)
end (*SD*)	29.30 (5.21)	32.10 (5.78)	29.45 (6.63)	29.45 (7.60)

Notes. DT discrimination training, *FB* feedback, *STAI* State-Trait-Anxiety-Inventory State Scale, *Negative Mood* and *Positive Mood* are subscales of the Positive and Negative Affect Schedule (PANAS).

6.1.2 Discrimination performance during training

Performance on discrimination training, i.e., percentage of mistakes, was analyzed with an ANOVA with within-subject factors learning (Part1, Part2) and comparison (i.e.,CS+ - CS+, CS+ -GS1, CS+ -GS2, CS+ -GS3, CS+ -GS4, CS+ -CS- in the relevant_DT group, or rather Line1-Line1, Line1-Line2, Line1-Line3, Line1-Line4, Line1-Line5, Line1-Line6 in the irrelevant_DT group) and the between-subjects factor fear relevance (relevant_DT, irrelevant_DT), but without factor feedback because of missing data.

ANOVA on discrimination performance during training showed a marginal Learning × Comparison interaction ($F(1.97, 74.72) = 2.74$, $p = .072$, $\eta_p^2 = .07$, **Supplementary Table 3**), suggesting that the ability to discriminate either the CS+ or the thin line from the other stimuli developed differently during the two parts of discrimination training. Post-hoc simple contrasts revealed a lower percentage of mistakes in Part 2 compared to Part 1 for only one comparison, comparing CS+ or Line1 to itself ($F(1, 38) = 6.88$, $p = .006$, $\eta_p^2 = .15$, all other p-values > .077, Bonferroni corrected $\alpha < .008$). I also found a main effect of comparison ($F(2.44, 92.62) = 39.79$, $p < .001$, $\eta_p^2 = .51$, **Supplementary Table 3**). Post-hoc simple contrasts (Bonferroni corrected $\alpha < .010$) showed that during the complete training participants made more false discriminations between CS+ or Line1 and CS+ or Line1 respectively ($F(1, 38) = 15.32$, $p < .001$, $\eta_p^2 = .29$) as well as between CS+ or Line1 and GS1 or Line2 respectively ($F(1, 38) = 65.83$, $p < .001$, $\eta_p^2 = .63$) as compared to between CS+ or Line1 and CS- or Line6 respectively. The percentage of mistakes for all other comparisons was not different from that for CS- or Line6 (all p-values > .067). In addition, the fear-relevant discrimination training had a lower percentage of mistakes compared to the fear-irrelevant alternative, as indicated by the main effect of fear

relevance ($F(1, 38) = 14.12$, $p < .001$, $\eta_p^2 = .27$, $M_{\text{fear-relevant}} = 5.80$, $SD_{\text{fear-relevant}} = 4.30$, $M_{\text{fear-irrelevant}} = 15.30$, $SD_{\text{fear-irrelevant}} = 10.96$). No other effect reached significance (all p-values > .229).

Supplementary Table 3 *Discrimination performance for each comparison of the training indicated by mistakes (percent)*

	CS+ - CS+ or L1 – L1	CS+ - GS1 or L1 – L2	CS+ - GS2 or L1 – L3	CS+ - GS3 or L1 – L4	CS+ - GS4 or L1 – L5	CS+ - CS- or L1 – L6
part 1	17.00	28.33	8.33	5.83	3.33	5.00
(SD)	(24.93)	(27.79)	(21.01)	(16.69)	(16.54)	(14.22)
part 2	6.00	37.50	5.00	3.33	0.83	1.67
(SD)	(8.10)	(35.56)	(14.22)	(12.63)	(5.27)	(7.36)
total	11.50	32.92	6.67	4.58	2.08	3.33
(SD)	(12.82)	(23.72)	(14.52)	(10.67)	(8.60)	(8.61)

Notes. L1 thinnest line, *L2-L5* lines with intermediate width, *L6* thickest line.

6.1.3 Stability of responses to conditioned stimuli throughout the experiment

I ran ANOVAs with the within-subject factors stimulus (CS+, CS-) and training (pre, post) to specifically examine the "stability" of fear learning during the generalization tests.

ANOVAs on CS responses during pre- and post-training generalization tests revealed significant main effects of stimulus for all dependent variables (US expectancy: $F(1, 79) = 325.43$, $p < .001$, $\eta_p^2 = .80$, CS+: $M = 57.88$, $SD = 28.20$, CS-: $M = 4.56$, $SD = 11.65$; arousal: $F(1, 79) = 251.48$, $p < .001$, $\eta_p^2 = .76$, CS+: $M = 6.10$, $SD =$

1.98, CS-: M = 2.26, SD = 1.48; valence: $F(1, 79) = 116.12$, $p < .001$, $\eta_p^2 = .60$, CS+: M = 3.55, SD = 1.52, CS-: M = 6.52, SD = 1.71; SCR: $F(1, 72) = 18.83$, $p < .001$, $\eta_p^2 = .21$; CS+: M = 0.04, SD = 0.05, CS-: M = 0.02, SD = 0.02), indicating stability of the conditioned responses during both generalization tests. The Stimulus × Training interaction was significant for US-expectancy ($F(1, 79) = 4.24$, $p = .043$, $\eta_p^2 = .05$, $CS+_{pre}$: M = 57.00, SD = 27.94, $CS-_{pre}$: M = 5.75, SD = 13.39, $CS+_{post}$: M = 58.75, SD = 28.61, $CS-_{post}$: M = 3.38, SD = 9.54) and trendwise for valence ratings ($F(1, 79) = 3.04$, $p = .085$, $\eta_p^2 = .04$, $CS+_{pre}$: M = 3.64, SD = 1.59, $CS-_{pre}$: M = 6.40, SD = 1.66, $CS+_{post}$: M = 3.46, SD = 1.46, $CS-_{post}$: M = 6.64, SD = 1.77). No other effects were found (all p values > .642). Post-hoc contrasts (Bonferroni corrected $\alpha < .012$) of these interactions suggest stable conditioned fear responses as CS+ vs. CS- ratings were higher both at pre (US expectancy: $F(1, 79) = 278.17$, $p < .001$, $\eta_p^2 = .78$, valence: $F(1, 79) = 84.25$, $p < .001$, $\eta_p^2 = .52$) and post test (US expectancy: $F(1, 79) = 305.73$, $p < .001$, $\eta_p^2 = .79$, valence: $F(1, 79) = 113.05$, $p < .001$, $\eta_p^2 = .59$). I also observed no significant changes in ratings for CS+ or CS- from pre- to post-training assessments (all p values > .184), except for a decrease in ratings of US expectancy to CS- from pre- to post-training assessment ($F(1, 79) = 8.8$, $p = .002$, $\eta_p^2 = .10$).

6.2 Study 1B

6.2.1 State questionnaires

As state control measures, the state version of the STAI and the Positive and Negative Affective Schedule (PANAS, Krohne, Egloff, Kohmann, & Tausch, 1996) were completed at the beginning and the end of the experimental protocol (**Supplementary Table 4** and **Supplementary Table 5**). The two non-discriminative training groups were compared using 2 (phase: beginning, end of experiment) × 2

(feedback: with, without) ANOVAs, which returned significant main effects of phase for all scales (**Supplementary Table 4**; state anxiety: $F(1, 38) = 23.40$, $p < .001$, $\eta_p^2 = .38$; negative mood: $F(1, 38) = 22.79$, $p < .001$, $\eta_p^2 = .37$) but positive mood ($F(1, 38) = 0.67$, $p = .417$, $\eta_p^2 = .02$) meaning that participants were more anxious and had higher negative mood, but no change of positive mood at the end of the experiment as compared to the beginning. All further effects including factor feedback did not reach significance (all p values > .623) indicating comparable emotional states in the two non-discriminative control training groups.

Supplementary Table 4 *Changes in the emotional state of the participants throughout the experimental procedure independently from feedback during non-discriminative control training*

	STAI state	*Negative Mood*	*Positive Mood*
begin (*SD*)	32.17 (6.08)	11.50 (2.18)	29.75 (5.98)
end (*SD*)	38.15 (7.87)	15.35 (5.56)	29.07 (6.24)

Notes. STAI State Trait Anxiety Inventory, *Negative Mood* and *Positive Mood* subscales of the Positive and Negative Affect Schedule (PANAS).

Additionally, I compared the emotional state of all training conditions with feedback using 2 (phase: beginning, end of experiment) × 2 (training: relevant_DT, irrelevant_DT, non-DT) ANOVAs, which returned significant main effects of phase for state anxiety ($F(1, 57) = 31.15$, $p < .001$, $\eta_p^2 = .35$) and negative mood ($F(1, 57) = 22.64$, $p < .001$, $\eta_p^2 = .28$). For positive mood, the main effect of phase reached marginal significance ($F(1, 57) = 3.46$, $p = .068$, $\eta_p^2 = .06$). These effects indicate that participants were more anxious and had higher negative mood as well as trendwise lower positive mood at the end of the experiment as compared to the beginning. Moreover, I observed a marginal significant interaction between phase and training for state anxiety ($F(2, 57) = 2.97$, $p = .059$, $\eta_p^2 = .09$; but no main effect of training: $F(2, 57) = 0.10$, $p = .905$, $\eta_p^2 <$

.01). This interaction reached significance for the negative mood scale ($F(2, 57) = 4.33$, $p = .018$, $\eta_p^2 = .13$) and was accompanied by a marginal significant main effect of training ($F(2, 57) = 2.67$, $p = .078$, $\eta_p^2 = .09$). No interaction effects were found for the positive mood scale (all p values > .109). Post-hoc simple contrasts (Bonferroni corrected $\alpha < .017$) for the interaction indicated that participants receiving the fear-irrelevant discrimination training or the non-discriminative training had higher state anxiety (irrelevant_DT: $F(1, 57) = 19.47$, $p < .001$, $\eta_p^2 = .25$; non-DT: $F(1, 57) = 16.06$, $p < .001$, $\eta_p^2 = .22$) and higher negative mood (irrelevant_DT: $F(1, 57) = 15.36$, $p < .001$, $\eta_p^2 = .21$; non-DT: $F(1, 57) = 15.82$, $p < .001$, $\eta_p^2 = .22$) at the end of the experiment as compared to the beginning, while the group who received the fear-relevant discrimination training showed no significant change of their state anxiety or mood throughout the experiment (all p values > .218).

Supplementary Table 5 *Changes in the emotional state of the participants throughout the experimental procedure dependent on the task*

	relevant_DT _FB	*irrelevant_DT _FB*	*non-DT _FB*	*non-DT _noFB*
STAI state				
begin (*SD*)	33.80 (5.83)	32.30 (6.00)	32.55 (6.57)	31.80 (5.70)
end (*SD*)	35.65 (5.62)	38.85 (10.48)	38.50 (9.12)	37.80 (6.62)
Negative Mood				
begin (*SD*)	11.35 (2.30)	11.35 (2.70)	11.90 (2.59)	11.10 (1.65)
end (*SD*)	11.65 (1.63)	14.75 (4.61)	15.35 (5.80)	15.35 (5.46)
Positive Mood				
begin (*SD*)	33.80 (5.28)	30.55 (7.06)	29.65 (5.06)	29.85 (6.91)
end (*SD*)	32.10 (5.78)	29.45 (7.60)	29.00 (5.79)	29.15 (6.82)

Notes. DT discrimination training, *FB* feedback, *STAI* State-Trait-Anxiety-Inventory State Scale, *Negative Mood* and *Positive Mood* are subscales of the Positive and Negative Affect Schedule (PANAS).

6.2.2 Discrimination performance during trainings

6.2.2.1 Discrimination vs. non-discriminative trainings

Training performance, i.e., percentage of mistakes, was analyzed with an ANOVA with within-subject factor learning (Part1, Part2) and between-subjects factor training (relevant_DT_FB, irrelevant_DT_FB, non-DT_FB).

The ANOVA showed a significant effect of learning ($F(1, 57) = 7.59$, $p = .008$, $\eta_p^2 = .12$; **Supplementary Table 6**), indicating that participants became better in their task, be it discriminative or non-discriminative. However, a main effect of training ($F(2,$

57) = 12.91, $p < .001$, $\eta_p^2 = .31$; **Supplementary Table 6**) suggests that the different training groups did not do equally well. Post-hoc simple contrasts (Bonferroni corrected $\alpha < .017$) showed that the fear-irrelevant discrimination-training group had a higher percentage of incorrect answers during the complete training than the fear-relevant discrimination-training group ($F(1, 57) = 14.88$, $p < .001$, $\eta_p^2 = .21$). The same was true in comparison to the non-discriminative group ($F(1, 57) = 22.96$, $p < .001$, $\eta_p^2 = .29$). The comparison of non-discriminative and fear-relevant discrimination-training groups resulted non-significant ($F(1, 57) = 0.87$, $p = .354$, $\eta_p^2 = .01$).

Supplementary Table 6 *Discrimination performance of discrimination trainings or non-discriminative control training indicated by mistakes (percent)*

	relevant_DT _FB	*irrelevant_DT _FB*	*non-DT _FB*	*non-DT _noFB*
part 1 (*SD*)	7.00 (5.01)	18.80 (18.94)	5.40 (11.18)	12.80 (22.87)
part 2 (*SD*)	4.60 (4.55)	11.80 (7.51)	1.60 (2.72)	4.40 (9.62)
total *SD)*	5.80 (4.87)	15.30 (14.66)	3.60 (8.26)	8.60 (17.83)

Notes. DT discrimination training, *FB* feedback.

6.2.2.2 Non-discriminative training with vs. without feedback

Training performance i.e., percentage of mistakes, during the non-discriminative training with or without feedback, was analyzed with an ANOVA with within-subject factor learning (Part1, Part2) and between-subjects factor feedback (with, without).

The ANOVA on performance showed a significant effect of learning ($F(1, 38) = 5.22$, $p = .028$, $\eta_p^2 = .12$; **Supplementary Table 6**), indicating that participants improved during the task independently from feedback. Moreover, the observed

improvement was not especially strong in the feedback group, as the Learning × Feedback interaction did not reach significance ($F(1, 38) = 0.74$, $p = .394$, $\eta_p^2 = .02$; **Supplementary Table 6**). The main effect of feedback was also non-significant ($F(1, 38) = 2.25$, $p = .142$, $\eta_p^2 = .06$), meaning that both groups also had a comparable overall performance.

6.3 Study 2

6.3.1 Questionnaires

The scores of the Anxiety Sensitivity Index (ASI), Agoraphobic Cognitions Questionnaire (ACQ) and the Social Phobia Anxiety Inventroy (SPAI), which are aggregated to one anxiety score in Study 2 are listed in **Supplementary Table 7**. Moreover, STAI Trait scores are indicated (**Supplementary Table 7**).

Supplementary Table 7 *Raw questionnaire scores of the four groups*

	total sample	*relevant_DT _noFB*	*relevant_DT _FB*	*irrelevant_DT _noFB*	*irrelevant_DT _FB*
ASI *(SD)*	14.80 (0.82)	15.58 (0.74)	13.28 (0.78)	16.53 (0.99)	14.09 (0.74)
ACQ *(SD)*	1.38 (0.28)	1.39 (0.25)	1.30 (0.22)	1.44 (0.32)	1.37 (0.33)
SPAI (*SD*)	1.62 (9.59)	1.67 (9.52)	8.85 (13.28)	1.69 (9.80)	1.52 (10.06)
STAI (*SD*)	35.62 (8.52)	35.55 (8.37)	34.94 (7.93)	38.17 (9.30)	34.06 (8.13)

Notes. DT discrimination training, *FB* feedback, *ASI* Anxiety Sensitivity Index 3, *ACQ* Agoraphobic Cognitions Questionnaire, *SPAI* Social Phobia and Anxiety Inventory, *STAI* Trait scale of the State Trait Anxiety Inventory.

As state control measures, the state version of the Spielberger State-Trait Anxiety Inventory (STAI; Dymond et al., 2015; Laux et al., 1981) and the Positive and Negative Affective Schedule (PANAS, Krohne, Egloff, Kohmann, & Tausch, 1996) were completed at the beginning and the end of the experimental protocol (**Supplementary Table 8** and **Supplementary Table 9**). The four groups were compared using 2 (phase: beginning, end of experiment) × 2 (fear relevance: relevant_DT, irrelevant_DT) × 2 (feedback: with, without) ANOVAs, which returned significant main effects of phase for all scales (**Supplementary Table 8**; state anxiety: $F(1, 235) = 137.10$, $p < .001$, $\eta_p^2 = .37$; negative mood: $F(1, 235) = 97.06$, $p < .001$, $\eta_p^2 = .29$; positive mood: $F(1, 235) = 86.81$, $p < .001$, $\eta_p^2 = .27$) meaning that participants were more anxious, had higher negative mood and lower positive mood at the end of the experiment as compared to the beginning. In addition, participants in the groups without feedback had higher positive affect throughout the experiment ($F(1, 235) = 5.80$, $p = .017$, $\eta_p^2 = .02$, **Supplementary Table 9**). All further effects including factor including feedback or fear relevance did not reach significance (all p values > .245) indicating a largely similar emotional state in all training groups.

Supplementary Table 8 *Changes in the emotional state of the participants throughout the experimental procedure independently from the task*

	STAI state	*Negative Mood*	*Positive Mood*
begin (*SD*)	33.04 (7.16)	11.73 (2.40)	31.00 (6.20)
end (*SD*)	40.45 (10.51)	15.27 (5.93)	27.83 (7.24)

Notes. STAI State Trait Anxiety Inventory, *Negative Mood* and *Positive Mood* subscales of the Positive and Negative Affect Schedule (PANAS).

Supplementary Table 9 *Changes in the emotional state of the participants throughout the experimental procedure dependent on the task*

	relevant_DT _noFB	*relevant_DT _FB*	*irrelevant_DT _noFB*	*irrelevant_DT _FB*
STAI state				
begin (*SD*)	31.41 (6.42)	33.30 (6.70)	33.91 (7.24)	33.51 (8.07)
end (*SD*)	39.11 (11.21)	41.25 (9.80)	41.56 (11.13)	39.78 (9.72)
Negative Mood				
begin (*SD*)	11.23 (1.61)	11.82 (2.36)	11.57 (2.07)	12.36 (3.24)
end (*SD*)	14.76 (5.88)	15.39 (5.74)	15.13 (5.53)	15.63 (6.65)
Positive Mood				
begin (*SD*)	32.64 (5.38)	29.86 (6.47)	31.18 (6.77)	30.20 (5.82)
end (*SD*)	29.69 (7.50)	27.23 (7.10)	27.98 (7.23)	26.27 (6.84)

Notes. DT discrimination training, *FB* feedback, *STAI* State-Trait-Anxiety-Inventory State Scale, *Negative Mood* and *Positive Mood* are subscales of the Positive and Negative Affect Schedule (PANAS).

6.3.2 Discrimination performance during training

The discrimination training performance, i.e., percentage of mistakes, first was analyzed with an ANOVA having the within-subject factors learning (Part1, Part2) as well as the between-subjects factors fear relevance (relevant_DT, irrelevant_DT) and feedback (with, without).

Then, discrimination training performance, was analyzed separately for fear-relevant and fear-irrelevant discrimination-training groups, including an additional within-subject factor comparison (i.e. CS+ - CS+, CS+ -GS1, CS+ -GS2, CS+ -GS3,

CS+ -GS4, CS+ -CS- in the relevant_DT group, or Line1-Line1, Line1-Line2, Line1-Line3, Line1-Line4, Line1-Line5, Line1-Line6 in the irrelevant_DT group). The ANOVAs further included the within-subject factor learning (Part1, Part2) and the between-subjects factor feedback (with, without).

The first ANOVA on discrimination performance during training showed significant main effects of learning ($F(1, 240) = 19.21$, $p < .001$, $\eta_p^2 = .07$; Part 1: $M = 11.74$, $SD = 12.54$), Part 2: $M = 8.54$, $SD = 8.18$), fear relevance ($F(1, 240) = 11.06$, $p = .001$, $\eta_p^2 = .04$; **Supplementary Table 10**) and feedback ($F(1, 240) = 13.34$, $p < .001$, $\eta_p^2 = .05$; **Supplementary Table 10**) indicating that discrimination performance improved during training in all groups, but was generally better in fear-relevant vs. fear-irrelevant, and feedback vs. no-feedback groups. None of the conditions showed a particularly strong improvement of perceptual discrimination, as no interaction including a between-subjects factor returned significant (all p values > .444).

Supplementary Table 10 *Discrimination performance of the four groups across all comparisons indicated by mistakes (percent)*

	relevant_DT _FB	*irrelevant_DT _FB*	*relevant_DT _noFB*	*irrelevant_DT _noFB*
part 1 (*SD*)	7.44 (10.13)	11.88 (8.87)	12.95 (14.11)	15.12 (15.51)
part 2 (*SD*)	4.81 (5.80)	8.79 (5.80)	8.44 (8.80)	12.41 (10.18)
total (*SD*)	6.12 (8.32)	10.33 (7.63)	10.69 (11.92)	13.76 (13.13)

Notes. DT discrimination training, *FB* feedback.

The ANOVA on discrimination performance of the fear-relevant training groups showed significant main effects of learning ($F(1, 117) = 8.34$, $p = .005$, $\eta_p^2 = .07$, Part 1: $M = 9.36$, $SD = 22.04$, Part 2: $M = 6.40$, $SD = 18.48$) and feedback ($F(1, 117) = 6.45$, p

= .012, η_p^2 = .05; fb: *M* = 5.95, *SD* = 16.69, no_fb: *M* = 10.11, *SD* = 23.79), confirming the main effects of the previous ANOVA. Furthermore, there was a significant main effect of comparison (*F*(1.74, 203.85) = 63.45, *p* < .001, η_p^2 = .35; see **Supplementary Table 11**). Post-hoc simple contrasts (Bonferroni corrected α < .010) showed that during the complete training the percentage of incorrect discriminations was higher for the two most difficult comparisons, i.e. CS+ - CS+ (*F*(1, 117) = 40.43, *p* < .001, η_p^2 = .26) and CS+ - GS1 (*F*(1, 117) = 85.67, *p* < .001, η_p^2 = .42), compared to the least difficult, i.e. CS+ - CS-. The percentage of mistakes of all other comparisons did not differ from that of the least difficult (all *p*-values > .132). No interaction effect of the ANOVA returned significant (all *p*-values > .198).

The ANOVA on discrimination performance of the fear-irrelevant training groups revealed a significant main effects of feedback (*F*(1, 123) = 5.41, *p* = .022, η_p^2 = .04, fb: *M* = 10.35, *SD* = 21.29, no_fb: *M* = 14.11, *SD* = 26.96), confirming less mistakes in perceptual discrimination in the feedback compared to the no-feedback group. Furthermore, there was a significant main effect of comparison (*F*(1.78, 218.68) = 187.53, *p* < .001, η_p^2 = .60; see **Supplementary Table 11**). Post-hoc simple contrasts (Bonferroni corrected α < .010) showed that during the complete training the percentage of incorrect discriminations was higher for the two most difficult comparisons, i.e. Line1 - Line1 (*F*(1, 123) = 68.56, *p* < .001, η_p^2 = .36) and Line1 - Line2 (*F*(1, 123) = 254.31, *p* < .001, η_p^2 = .67), compared to the least difficult, i.e. Line1 - Line6. The percentage of mistakes of all other demand levels did not differ from that of the least difficult (all *p*-values > .020). In the fear-irrelevant training condition, the capacity to discriminate the thin line from the other stimuli changed differently during the two training parts, as indicated by a significant Learning × Comparison interaction (*F*(2.30, 283.26) = 7.67, *p* < .001, η_p^2 = .06). Post-hoc simple contrasts revealed a lower

percentage of mistakes in Part 2 compared to Part 1 for one comparison only, namely the comparison of Line1 with itself ($F(1, 123) = 16.89$, $p < .001$, $\eta_p^2 = .12$, all further p-values > .016, Bonferroni corrected $\alpha < .008$). No further effect of the ANOVA returned significant (all p values > .075).

Supplementary Table 11 *Discrimination performance for each comparison of fear-relevant or fear-irrelevant discrimination training indicated by mistakes (percent)*

CS+ - CS+	CS+ - GS1	CS+ - GS2	CS+ - GS3	CS+ - GS4	CS+ - CS-
9.16	28.01	3.78	2.10	2.10	2.10
(14.59)	(33.96)	(15.30)	(11.47)	(11.47)	(10.62)
L1 – L1	L1 – L2	L1 – L3	L1 – L4	L1 – L5	L1 – L6
11.52	45.07	6.13	2.80	3.73	3.47
(14.84)	(34.65)	(16.57)	(11.02)	(12.83)	(13.57)

Notes. L1 thinnest line, *L2-L5* lines with intermediate width, *L6* thickest line, indicated are means (with *SD*).

Publication list

Peer-reviewed journal articles

Imholze, C.*, **Hutterer, K.***, Gall, D., Dannlowski, U., Domschke, K., Leehr, E.J., . . . Gamer, M. (in press). Prediction of changes in negative affect during the COVID-19 pandemic by experimental fear conditioning and generalization measures – a longitudinal study. *Zeitschrift für Psychologie.*

*these authors contributed equally

Herzog, K., Andreatta, M., Schneider, K., Schiele, M. A., Domschke, K., Romanos, M., . . . Pauli, P. (2021). Reducing Generalization of Conditioned Fear: Beneficial Impact of Fear Relevance and Feedback in Discrimination Training. *Front Psychol, 12*(1648), 665711. doi:10.3389/fpsyg.2021.665711

Andreatta, M., Neueder, D., **Herzog, K.**, Genheimer, H., Schiele, M. A., Deckert, J., . . . Pauli, P. (2020). Generalization of Conditioned Contextual Anxiety and the Modulatory Effects of Anxiety Sensitivity. *Neurotherapeutics, 17*(3), 1239-1252. doi:10.1007/s13311-020-00831-8

Schiele, M. A., **Herzog, K.**,Kollert, L., Bohnlein, J., Repple, J., Rosenkranz, K., . . . Domschke, K. (2020). Affective temperaments (TEMPS-A) in panic disorder and healthy probands: Genetic modulation by 5-HTT variation. *World J Biol Psychiatry, 21*(10), 790-796. doi:10.1080/15622975.2019.1705999

Schiele, M. A., **Herzog, K**., Kollert, L., Schartner, C., Leehr, E. J., Bohnlein, J., . . . Domschke, K. (2020). Extending the vulnerability-stress model of mental disorders: three-dimensional NPSR1 x environment x coping interaction study in anxiety. *Br J Psychiatry, 217*(5), 645-650. doi:10.1192/bjp.2020.73

Contributions to international conferences

Klingelhöfer-Jens, M., **Hutterer, K.**, Schiele, M. A., Schümann, D., Deckert, J., Domschke, K.,… Lonsdorf, T. B. (2022). From then to now: Effects of childhood maltreatment on acquisition and generalization of conditioned fear.

Herzog, K., Andreatta, M., Schiele, M. A., Domschke, K., Deckert, J., & Pauli, P. (2021). How stable can fear generalization be?" *European Meeting on Human Fear Conditioning (EMHFC@home)*, Bochum, Germany.

Herzog, K., Andreatta, M., Domschke, K., Deckert, J., & Pauli, P. (2019). Reducing Generalization of Conditioned Fear: comparison of a fear-specific discrimination training with fear-unrelated control tasks. *World Association for Stress Related and Anxiety Disorders (WASAD)*, Würzburg, Germany.

Herzog, K., Andreatta, M., Domschke, K., Deckert, J., & Pauli, P. (2019). Reducing Generalization of Conditioned Fear: comparison of a fear-specific discrimination training with fear-unrelated training. *European Meeting on Human Fear Conditioning (EMHFC)*, Würzburg, Germany.

Herzog, K., Andreatta, M., Domschke, K., & Pauli, P. (2018). De-generalization of conditioned fear: feedback increases beneficial effects of a discrimination training. *Psychologie und Gehirn (PuG)*, Gießen, Germany.

Herzog, K., Andreatta, M., Domschke, K., & Pauli, P. (2018). Reversibility of generalized fear: comparison of a discrimination training with and without feedback. *European Meeting on Human Fear Conditioning (EMHFC)*, Cardiff, UK.

Herzog, K., Andreatta, M., Markfelder, S., Domschke, K., & Pauli, P. (2017). Can fear generalization be reversed by a discrimination training? *World Association for Stress Related and Anxiety Disorders (WASAD)*, Würzburg, Germany.

Herzog, K., Andreatta, M., Markfelder, S., Domschke, K., & Pauli, P. (2017). Can fear generalization be reversed by a discrimination training? *European Meeting on Human Fear Conditioning (EMHFC)*, Hamburg, Germany.

Curriculum Vitae

Affidavit

I hereby confirm that my thesis entitled “Treatment-like use of discrimination training to reduce generalization of conditioned fear“ is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date	Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation „Behandlungsähnlicher Einsatz eines Diskriminationstrainings zur Verringerung von Generalisierung konditionierter Furcht“ eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum	Unterschrift