**Fear Conditioning Biases in Anxiety Disorders: A Matter of Interpretation?**

Sara Scheveneels & Yannick Boddez

\_\_\_\_\_\_\_\_\_\_\_\_\_\_
Sara Scheveneels
sara.scheveneels@kuleuven.be

With a lifetime prevalence between 25 and 30 percent, anxiety disorders are a prevalent and debilitating condition, associated with significant individual and societal burden (Kessler et al., 2005). Research about factors involved in the onset, maintenance, and exacerbation of anxiety disorders has focused on the role of information selection and processing (Beck & Clark, 1997; Lau & Waters, 2016). For example, relative to non-anxious controls, patients diagnosed with an anxiety disorder show an attentional bias towards threat, including facilitated threat detection, difficulties in disengagement from threat and attentional avoidance (Bar-Haim et al., 2007; Cisler & Koster, 2010). Moreover, patients diagnosed with an anxiety disorder exhibit a tendency to consistently resolve ambiguous cues and situations in a specific direction by selecting threatening instead of benign meanings (Hirsch et al., 2016; and for an overview of the role of interpretation biases in emotional psychopathology, see chapter 2.1 by Woud & Hofmann). Another example concerns research that focuses on biases in fear conditioning. In this chapter, we will start with providing an overview of such fear conditioning biases in anxiety and discuss how these can be linked to interpretation biases. Subsequently, we will discuss different ways of defining biases. We will conclude this chapter with discussing how (fear conditioning) biases play a role in interventions directed at reducing anxiety.

1.7.1 Biases of Anxiety Patients in the Context of Fear Conditioning and Links to Interpretational Processing

Learning mechanisms play a crucial role in the etiology and maintenance of anxiety (Mineka & Zinbarg, 2006; Scheveneels et al., 2019). These learning mechanisms are often studied under highly-controlled conditions using (fear) conditioning procedures (Beckers et al., 2013; Lonsdorf et al., 2017). In a standard fear conditioning paradigm, a neutral stimulus (e.g., a geometrical figure) is paired with another stimulus that is aversive in nature (e.g., an electric shock). As a result of these stimulus pairings, responding indicative of fear and anxiety is elicited by the first (initially neutral) stimulus. This stimulus - to which responding changes conditional upon being paired with the aversive stimulus - is called a *conditional stimulus* or CS. The (aversive) stimulus that is paired with the CS and changes responding to it is called the *unconditional stimulus* (US). The response to the CS that changes due to CS-US pairings is termed the *conditional response* (CR). In the context of anxiety, amongst the responses of interest are US-expectancies (Boddez et al., 2013), subjective fear, physiological indices of anxiety (e.g., skin-conductance, fear-potentiated startle), and avoidance.

Interestingly, empirical evidence indicates that patients diagnosed with an anxiety disorder behave differently in fear conditioning procedures as compared to non-anxious controls. In particular, patients show a tendency or bias to respond fearfully to ambiguous or in fact safe stimuli in the context of fear conditioning. Interestingly, these observable differences in fear conditioning can be linked to interpretation biases. In this chapter we will describe how a tendency to interpret ambiguous or safe stimuli as threatening might (partially) mediate responding in a fear conditioning procedure[[1]](#footnote-1). We will start with providing a summary of the evidence for differences between anxiety patients and controls in different fear conditioning procedures (Figure 1.7.1), and then describe the relevance of biased interpretational processes during these conditioning procedures.

When adding a second stimulus that is never paired with the US (i.e., a CS-) to a standard fear conditioning procedure, it is observed that individuals diagnosed with an anxiety disorder tend to respond with increased anxiety or fear to this stimulus – a stimulus that in fact can be considered as safe – as compared to control participants (Duits et al., 2015; Dvir et al., 2019). Related to this, anxiety patients tend to show *reduced CS+/CS- discrimination* in fear conditioning procedures (Cooper et al., 2018). These observations (i.e., increased CS- responding and reduced discrimination) can be (partially) mediated by interpretation biases, namely the tendency of patients to misinterpret the safe CS- as threatening. A real-life translation of such reduced discrimination between danger and safety situations can, for example, be observed in the context of panic disorder, when innocuous physiological symptoms are misinterpreted and treated as a sign of having a heart attack (Haddad et al., 2012). Strongly related to this possible bias in responding fearfully to *safe* situations, individuals diagnosed with an anxiety disorder tend to react fearfully to *ambiguous* stimuli. We now discuss two fear conditioning procedures that induce an ambiguous situation.

First, individuals diagnosed with an anxiety disorder, as compared to non-anxious controls, are slower to learn that a threatening stimulus has become safe. This has been demonstrated using extinction procedures, in which a CS+ that was previously paired with a US is presented without the US (Hermans et al., 2006). In particular, *reduced and/or delayed extinction* is observed (i.e., stronger responding to the CS+ and / or slower reduction in responding) in patients compared to controls (Duits et al., 2015, 2017; Dvir et al., 2019)[[2]](#footnote-2). These results suggest that individuals diagnosed with an anxiety disorder need more learning experiences to acquire a sense of safety. Again, this can (partially) be driven by an interpretation bias in patients. Namely, the extinguished CS+ might still be interpreted as threatening despite experiences in which the US was absent. Possibly, patients may need more correcting experiences to change their interpretation of the CS+ from dangerous to safe, resulting in sustained CS+ responding during extinction. Translated to real life, after a car accident, fear will decrease with a slower rate in the case of new safe experiences with driving in individuals at risk of developing pathological anxiety. Linked to interpretation biases, these at risk individuals may continue to interpret driving a car as a dangerous, despite experiences with driving without accidents. Put differently, interpretation biases may hamper the expression of new learning experiences.

These extinction trials can be considered as instances of ambiguous situations. It has been argued that during extinction, the CS acquires an ambiguous meaning by not being paired with the US anymore. This is supported by extensive evidence on reinstatement and renewal of fear, showing that fear responding to the CS after extinction is modulated by context (Bouton, 1988, 2002). When an extinguished CS is presented in a threatening context again, fear can be ‘reinstated’, and when the CS is presented in another context than the extinction context, renewal of fear occurs. In order to explain this, it has been suggested that a CS that underwent extinction training has two meanings: a threatening and a safe one (due to acquisition training and the subsequent extinction training, respectively). Individuals diagnosed with anxiety disorders seem to have a tendency to still select the threatening meaning, which might be (partially) driven by a biased interpretation of the ambiguous CS as being dangerous (but see: McLaughlin et al., 2015).

A second observation is that anxiety patients tend to respond fearfully to (generalization) stimuli that are situated on the continuum between safe and threat signals. In fear conditioning procedures, a generalization test phase can be added after an acquisition phase with CS-US pairings. In this test phase, generalization stimuli are presented that are conceptually or perceptually related to the CS, and thus can be considered as (partly) ambiguous cues. Compared to controls, individuals diagnosed with anxiety disorders have shown more shallow generalization gradients, exhibiting stronger anxiety responding to stimuli that are dissimilar from the CS+ (El-Bar et al., 2017; Kaczkurkin et al., 2017; Lissek et al., 2010, 2014). This pattern is often referred to as overgeneralization. Overgeneralization can be affected by interpretational processes as well, namely the (ambiguous) generalization stimuli are misinterpreted as threatening, and as a result elicit elevated fear responding. Overgeneralization of fear responding can, for instance, occur when after being bitten by a dog, fear is not confined to this particular dog but spreads to other dogs that look similar or even to other animals that can bite such as cats or rabbits. It is, however, important to note that patient-control differences regarding overgeneralization are not always unequivocally replicated (Ahrens et al., 2016; Morey et al., 2020; Tinoco-González et al., 2015).

**Figure 1.7.1***Overview of some fear condition procedures*



In extension to this, a *covariation or expectancy bias* is observed in patients in the context of fear conditioning (Tomarken et al., 1989; Wiemer & Pauli, 2016). In ambiguous situations, anxiety patients tend to overestimate the contingency between fear-relevant stimuli (CSs) and aversive consequences (USs) despite the actual absence of a correlation (i.e., illusory correlation). Evidence for this covariation bias has been found in a variety of fears and anxiety disorders, including social anxiety (De Jong et al., 1998), specific fears (de Jong et al., 1992, 1995a; Pauli et al., 1998), and panic disorder (Pauli et al., 2001). For example, in an experimental context, it has been demonstrated that individuals with spider phobia overestimate the covariation between a spider picture and an electric shock (de Jong et al., 1992), and individuals retaining this bias after treatment are more vulnerable to relapse (De Jong et al., 1995b). These exaggerated expectations of an aversive outcome (US) after a fear-relevant stimulus (CS) can even occur prior to any actual conditioning experience or stimulus pairings, also termed *a priori expectancy bias* (Van Overveld et al., 2010; Wiedemann et al., 2001). Notably, this a priori expectancy bias can still be due to real-life, naturalistic learning experiences outside the experimental context. For example, a spider phobic who had a panic attack during a confrontation with a spider (outside the laboratory) might infer from this experience that the chances that the picture of the spider in the experiment will also be paired with something unpleasant are rather high, despite the lack of an actual contingency between the picture of the spider and the shock in the experimental task.

In conclusion, anxiety patients tend to exhibit biases in fear responding in conditioning tasks, and such biases may be partly linked to biases in interpretational processing, e.g., in how someone is interpreting a stimulus in a conditioning procedure. Regarding biases in fear responding, it has been claimed that these biases are not merely epiphenomena, consequences of anxiety disorders or just a diagnostic marker. Impaired discrimination, reduced extinction, and overgeneralization have also been argued to be involved in the onset of anxiety symptoms (Scheveneels et al., 2021). Prospective studies have been particularly influential in this regard. For example, Lommen et al. (2013) found in a sample of soldiers that reduced extinction learning pre-trauma (i.e., before deployment in Afghanistan) predicted PTSD symptom severity after deployment. The effect remained while controlling for pre-trauma PTSD symptoms and other risk factors such as neuroticism[[3]](#footnote-3). Based on the results of this and other prospective studies (Lenaert et al., 2014; Lommen et al., 2013; Orr et al., 2012; Guthrie & Bryant, 2006; Scheveneels et al., 2021), it has been argued that fear conditioning biases can constitute a predisposing factor that make individuals more vulnerable for developing an anxiety disorder (Britton et al., 2011). In addition, fear conditioning biases may contribute to the *persistence* of anxiety. For example, Sijbrandij et al. (2013) showed that reduced safety learning in a fear conditioning procedure in soldiers approximately 2 months after their deployment to Afghanistan predicted the persistence of post-traumatic stress symptoms at 9 months after deployment. To the extent that the fear conditioning biases are driven by skewed interpretations at the mental level, this suggests the important role that interpretational processes may play in the onset of pathological anxiety.

In this section, we provided a summary of differential behavioral outcomes in fear conditioning procedures (i.e., termed fear conditioning biases) in individuals diagnosed with an anxiety disorder and linked these to biases in interpretational processing (as a latent mental process). In the next section, we further elaborate on how these biases can be defined and theoretically linked to each other.

1.7.2 Defining Biases: Bias as (Observable) Behavior and Bias as a (Cognitive) Deficit

In this section we consider different perspectives on biases, namely biases as an effect (observable behavior) and biases as a(n) (underlying) mental process or deficit. We apply this perspective to both fear conditioning and interpretation biases. However, it should be noted that this perspective is not confined to these specific biases and can also be extended to other types of bias (De Houwer, 2019). We discuss how etiological factors including genetic and temperamental factors, in interaction with learning history, could give rise to biased behaviors (in fear conditioning tasks). This relation can be fully or partially mediated by unobservable latent mental processes (i.e., a cognitive deficit). A visual representation of this model is displayed in Figure 1.7.2.

**Figure 1.7.2**

*Mediational model of biases in behavior and its impact on anxiety, mediated by latent mental processes*

1.7.2.1 Bias as (Observable) Behavior

In line with how we described fear conditioning biases in the previous section, biases can be primarily defined in terms of *behavior* that is *observed in individuals diagnosed with an anxiety disorder* (in fear conditioning tasks) and more specifically, how this behavior deviates from what healthy controls would do (Scherer, 2020). In particular, in the context of fear conditioning, patients tend to exhibit stronger fear responding to safe or ambiguous stimuli or situations than controls. Biases are then defined in terms of *observable behavior*, in this case in a fear conditioning task. This definition can be extended to other biases as well, such as differences in performance of patients as compared to non-anxious controls in a dot-probe task or in a visual search task in the case of attention bias (Cisler et al., 2009). Similarly, with regard to interpretation bias, lexical decision tasks, sentence completion tasks, or scrambled sentences tasks have been used (amongst other tasks) to demonstrate differences between anxiety patients and controls (Schoth & Liossi, 2017; Würtz et al., 2022; and for an overview of measures to assess interpretation biases, see chapter 1.3 by Würtz & Sanchez-Lopez). It can be noted that this definition of interpretation bias as observable behavior (or effect) differs from how we described interpretation bias previously (cf. section 1) as a latent mental process that (partially) explains the observable behavior in fear conditioning tasks (see later).

1.7.2.2 The Etiology of (Fear Conditioning) Biases

In going one step further than merely inspecting observable behavior, we can try to explain these observable differences between patients and controls. This concerns the *etiology* of biases or the question: “Why do some people show biased behavior whereas others do not?”. This is a relevant question since it has been argued that biased behavior (observed in fear conditioning) is not just a consequence of anxiety disorders, but that it is also observed in healthy individuals with an increased vulnerability to develop an anxiety disorder (Scheveneels et al., 2021).

1.7.2.2.1 Temperamental and Genetic Factors

A first possibility is that some individuals are born with a bias. In line with this view, temperamental factors and genetic variations have been linked to biased responding in fear conditioning procedures (Lonsdorf & Kalisch, 2011; Lonsdorf & Merz, 2017). With regard to temperamental factors, individual differences in *neuroticism* (i.e., the tendency to express negative affect) and *trait anxiety* have been examined amongst others temperamental factors. Studies on the association between individual differences in fear conditioning on the one hand and neuroticism and trait anxiety on the other hand provide a heterogeneous picture (Lonsdorf & Merz, 2017). The majority of studies report null findings regarding the association with (differential) acquisition learning for neuroticism (e.g., Arnaudova et al., 2017; Lommen et al., 2010; Tzschoppe et al., 2014; Torrents-Rodas et al., 2013), and a large number of studies finds null results for trait anxiety as well (e.g., Arnaudova et al., 2013; Sehlmeyer et al., 2011). Regarding generalization, results are mixed (e.g., Arnaudova et al., 2017; Lommen et al., 2010; Torrents-Rodas et al., 2013), with a recent meta-analysis revealing a small, positive relationship between anxious personality (including neuroticism and trait anxiety) and fear generalization (Sep et al., 2019). For fear extinction, many studies fail to find an association with neuroticism (e.g., Pineles et al., 2009; Rattel et al., 2020; Tzschoppe et al., 2014) and for trait anxiety findings are inconsistent with some studies showing deficient extinction learning in high trait anxious individuals but only in some of the outcome measures (Barrett & Armony, 2009; Gazendam et al., 2013; Sehlmeyer et al., 2011). Regarding the link with interpretation bias, one possibility is that individuals high on neuroticism/trait anxiety exhibit a stronger tendency to select a threatening meaning, which might give rise to stronger fear responding in a fear conditioning task (Lommen et al., 2010; Salemink & van den Hout, 2010).

Studies uncovering genetic risk factors have revealed that variation in the serotonin transporter (5HTT) and the catechol-o-methyltransferase (COMT) genes are associated with reduced CS+/CS- discrimination, CS- responding (i.e., less CS- inhibition), and impaired extinction (i.e., less CS- inhibition and stronger CS+ responding) (Garpenstrand et al., 2001; Lonsdorf et al., 2009). At the same time, these genetic polymorphisms explain a relatively small proportion of the total variance in fear conditioning behavior, suggesting that this at best does not tell the whole story (for an overview of genetic factors underlying interpretation biases, see chapter 1.8 by Vincent & Fox).

1.7.2.2.2 Environmental Factors and Learning History

Additionally and / or alternatively, environmental factors such as learning experiences might explain why some individuals develop a bias and others do not. Anxiety patients and individuals vulnerable to develop an anxiety disorder might show a bias in fear conditioning tasks due to a *different learning history* than non-anxious controls. This can include, for instance, dysfunctional interpersonal experiences (e.g., family dynamics, attachment) or stressful life events (e.g., trauma). For example, an individual growing up in an unpredictable environment (e.g., with an emotionally volatile father) may have learned that safety on one or even several occasions (e.g., father behaving friendly and sweet) does not imply that the situation is always safe (e.g., as the father could suddenly and unpredictably behave aggressively). Such individuals may have learned to be on guard even in seemingly safe situations and therefore remain on guard when confronted with an apparently safe stimulus in a fear conditioning task (CS- or extinguished CS). This would result in maintained fear responding towards such stimuli despite ‘safe experiences’ (Spix et al., 2021; Vervliet & Boddez, 2020). Moreover, if also other, similar individuals (e.g., a male teacher of a similar age as father, an uncle) behaved in an unpredictable, volatile manner, the individual may infer that similar people or situations pose similar risks. That is, one may learn to generalize. As a consequence, anxiety responding might generalize strongly in a generalization test in a fear conditioning procedure (Boddez et al., 2017). These learning experiences and environmental factors may then interact with biological and predisposing factors in explaining biases in current behavior (Lonsdorf & Merz, 2017). Whereas research consistently revealed associations between learning history such as dysfunctional interpersonal experiences and stressful life events on the one hand, and an increased risk for developing pathological anxiety on the other hand (e.g., Jinyao et al., 2012; Lindert et al., 2013; Yap et al., 2014), the amount of studies that directly links learning history and fear conditioning biases is scarce. Linking this to interpretation bias, it can be hypothesized that the behavior of individuals with a particular learning history (cf. above) in a fear conditioning task is (partially) driven by interpretational processes, for instance, as a result of these learning experiences these individuals might show a tendency to misinterpret safe situations as threatening.

1.7.2.3 Bias as a Latent Mental Process

So far we discussed biases as observable behavior resulting from the interaction between predispositional biological factors and learning history[[4]](#footnote-4). Additionally, it is often assumed that this relation is mediated by particular *latent mental processes* (Figure 1.7.2; also see: De Houwer, 2011). In particular, learning history in interaction with an inborn vulnerability could generate a particular internal model of the world, which can take the form of cognitive schemata, processing styles, trust issues, expectancy bias, and so on. On their turn, these mental processes would drive the behavior in experimental tasks such as fear conditioning tasks, dot-probe tasks, and lexical decision tasks.

Some researchers do not put the mental processes in this merely mediational role, but see them as a unique and independent source of variance in (biased) behavior. From this perspective, individuals diagnosed with an anxiety disorder do not (only) differ in their learning history and inborn vulnerability, but also in the way in which they cognitively process information. In line with this, biases have been ascribed as deficits or as ‘flaws in the design of the mind’ (Haselton et al., 2015). This deficit view readily aligns with terminology suggesting that the patient *is biased* or *has a bias* (De Houwer, 2019). In contrast, perspectives that put more emphasis on environmental determinants, including some learning perspectives, align more readily with terminology suggesting that prior events *bias the behavior* of the patient. Relying on the examples introduced above, one could then say that prior exposure to an emotionally volatile father biases (i.e., systematically impacts) fear extinction or that exposure to a CS+ *biases responding* to a CS- or to a GS. That is, the biasing factors are (also) situated in the environment rather than (only) in the mind of the patient (see De Houwer, 2019).

We now describe one influential mental process that accounts for biases in fear conditioning, suggesting that patients would mentally rely on a *better safe than sorry processing strategy* (Lommen et al., 2010; Van den Bergh et al., 2021). Recently, Van den Bergh and colleagues (2021) further elaborated on this account by applying a predictive-processing approach. This approach states that individuals have prior beliefs or expectancies about the world (e.g., one might expect to get a shock, to be socially rejected, to be bitten by a dog). Learning experiences, combined with biological factors, may trigger such prior beliefs (cf. above). These prior beliefs or expectancies, however, may differ from the actual outcome in a given situation (e.g., getting no shock, being accepted by others, not being bitten). In case of such a mismatch between the expected and real outcome, a prediction error occurs. The brain supposedly functions as a prediction machine that aims to minimize these prediction errors. If, however, such mismatch occurs, this can be handled in several ways. The most straightforward way is to update prior beliefs and expectancies in a way that they better fit with reality (Rescorla & Wagner, 1972). Another way of dealing with the mismatch is by responding in a perceptual, behavioral, and physiological way that generates input that is consistent with the prior expectancies. This then results in discarding the information from the environment (e.g., offered during the experiment or experiences), i.e., information that does not match. Transferring this to the context of anxiety, individuals vulnerable for anxiety disorders (and broader psychopathology) might rely on information processing that tends to be low in sensory-perceptual detail (i.e., oversimplifying input from the environment) and heavily relies on categorical threat-related priors or beliefs (i.e., *jumping to conclusions*). In fact, such processing styles can also be related to biased interpretational processing, e.g., in the context of *jumping to conclusions*, whereby unwarranted assumptions are made based on limited information (e.g., a threatening interpretation is generated very quickly). As a consequence, processing is then primarily informed by threat-related priors or beliefs (e.g., others don’t like me and reject me) at the cost of actual (disconfirming) input (e.g., others might be interested in me). The benefit of such processing strategy is that it allows for greater speed in categorizing input and may reduce uncertainty in the short term. At the same time, this processing heuristic is believed to result in poor updating of prior threat-related beliefs, and thus allow these beliefs to remain dominant (stagnated error-reduction process). This way, prior threat-related beliefs can be maintained despite their mismatch with reality. In the context of fear conditioning, this could explain the link between fear conditioning and interpretation biases, i.e., prior threat-related beliefs may be associated with a biased interpretational processing style, which, in turn, can explain the observation that anxiety patients tend to interpret in fact safe stimuli or contexts as being threating, despite having safe experiences with the stimulus/context (e.g., a CS- or extinguished CS). It would be predicted that patients experience fewer prediction errors and are less sensitive to corrective experiences.

The mediational model that we introduced earlier (Figure 1.7.2) can also be useful to understand the potential link between interpretation bias (as a mental process) and biases in a fear conditioning task as observable behavior (see Figure 1.7.3)[[5]](#footnote-5). Earlier we described how differences in behavior in fear conditioning between patients and controls might (partially) be driven by interpretation bias. Interpretation bias can be seen as an inborn and learned processing style or underlying cognitive tendency that reflects a tendency to interpret unclear or ambiguous situations or stimuli as threatening. In fear conditioning tasks, participants are typically presented with ambiguous situations, for example stimuli that show some resemblance to the threatening stimulus or CS (i.e., generalization test) or stimuli that have acquired both a threatening and a safe meaning (i.e., extinction procedure). In line with this, it has been argued that the more ambiguous and unclear the situation (also referred to as ‘weak situations’), the more sensitive the procedure is to pick up differences between individuals vulnerable for anxiety and controls (Beckers et al., 2013; Lissek et al., 2006). Hence, in these weak or ambiguous situations, there might be more room for interpretation biases (as a mental process) to exert an influence on fear responding in the conditioning task (i.e., the observable behavior). For example, a generalization or extinction stimulus that is ambiguous and is interpreted as threatening will then elicit stronger fear responding.

**Figure 1.7.3**

*Mediational model of the link between interpretation bias as a mental process and biased behavior in fear conditioning tasks*



In conclusion, we discussed different perspectives on biases and integrated these in a mediational model (Figures 1.7.2 and 1.7.3). We started with defining a bias in terms of observable behavior (for instance in experimental tasks such as a fear conditioning task). It was then discussed how an interaction between biological (e.g., genetic and temperamental) factors and learning history could lie at the basis of biased behavior. Latent mental information processing strategies, such as a *better safe than sorry processing strategy* or interpretation biases, if defined as a mental process, could then mediate the relation between etiological factors and observable biased behavior (in this case in fear condition tasks).

1.7.3 The Role of (Fear Conditioning) Biases in Reducing Anxiety

In this section, we discuss how fear conditioning biases might play a role in the treatment of anxiety and how knowledge about fear conditioning biases, including the potential role of interpretation biases, may inform and enhance treatment strategies.

Exposure therapy is the treatment of choice for anxiety disorders and has proven highly effective for, for example, social anxiety disorder (Mayo-Wilson et al., 2014; Scaini et al., 2016), specific phobias (Wolitzky-Taylor et al., 2008), posttraumatic stress disorder (Cusack et al., 2016), and panic disorder (Pompoli et al., 2018). During exposure therapy, the patient is confronted with fear-eliciting stimuli that they would normally avoid. One of the currently prevailing models of exposure therapy is the inhibitory learning theory (ILT; Craske et al., 2008, 2014, 2022). Based on associative learning models and extinction research, this theory prescribes that expectancy violation or prediction error-violation are crucial in driving the effects of exposure through the formation of an inhibitory association that counteracts the original fear association (Boddez et al., 2020; Bouton, 1988, 2002; Rescorla & Wagner, 1972). More specifically, if a patient experiences during exposure that the fear eliciting stimulus (or CS) is not paired with the expected aversive outcome (or US), their fearful expectancy would be violated and the inhibitory association would gain strength. A stronger inhibitory association, supposedly, would be reflected in less feared expectancies (but see Boddez et al., 2020), representing a reduction in fearful cognitions after exposure therapy.

Based on these theoretical premises, it can be predicted that individuals who show poor inhibitory or safety learning would show a reduced outcome of exposure treatment. This has (at least partially) been confirmed in several studies on predicting exposure therapy outcome by laboratory-based extinction learning (Forcadell et al., 2017; Lange et al., 2020; Raeder et al., 2020; Scheveneels et al., 2021). However, significant results are typically found in only a subset of the dependent variables and are not always replicated by other studies, with some studies finding no association (Wannemueller et al., 2018) and others showing different associations than what would be expected (Geller et al., 2019). An association between reduced safety learning and worse exposure therapy outcome would also be in line with the predictive processing account and *better safe than sorry processing* strategy. Due to stagnated error reduction processes some individuals might be less sensitive for corrective experiences and would be impaired in updating their fearful expectancies or prior beliefs.

Impairments in safety learning could have several implications for the treatment of anxiety. First, individuals with poor safety learning might need more corrective experiences to adjust both their fearful expectancies and their biased interpretation of the (previously) feared stimulus. In addition, a stronger mismatch between their fearful expectancies and the actual outcome might be required to counteract impairments in inhibitory or safety learning. In line with this, several strategies have been proposed to maximize expectancy violation and inhibitory learning during exposure therapy (Craske et al., 2014; Pittig et al., 2018; Weisman & Rodebaugh, 2018). Some evidence has been found for the prediction that expectancy violation is related to better exposure treatment outcome (Deacon et al., 2013). However, more research is needed to further test the effects of these strategies (Craske et al., 2022).

Departing from a predictive-processing framework, it has been emphasized that treatment interventions should (additionally) target the way threat-relevant information is processed at all levels, including defensive action tendencies that are part of the fearful prior beliefs or expectancies (Van den Bergh et al., 2021). This implies encouraging openness to process threat-relevant information in a more detailed and less biased way and disengage from defensive-action tendencies during processing. It has been claimed that more detailed sensory processing will then promote updating the fearful prior expectancies. In other words, not only learning that the aversive stimulus or US is not coming will be a target in effective intervention, but also encouraging information processing that is no longer inspired by a *better safe than sorry strategy* but by a *wait-and-see attitude* (Van den Bergh et al., 2021).

In line with the above, it can be predicted that individuals with a stronger interpretation bias might also be less sensitive to the corrective experiences in exposure therapy and CBT because they are expected to interpret unclear or ambiguous outcomes as threatening (but see Baumgardner et al., 2022, in which baseline interpretation bias did not predict anxiety levels after CBT, or Mobach et al., 2021, in which stronger baseline interpretation bias predicted better CBT outcome). Moreover, changing interpretation bias (e.g., by cognitive bias modification techniques) could potentially enhance the sensitivity to corrective experiences and the efficacy of exposure treatment or CBT (Butler et al., 2015). However, a study by Steinman and Teachman (2014) found no evidence for an added effect of cognitive bias modification on exposure therapy. Some research suggests that interpretation bias decreases after cognitive-behavioral therapy (Baumgardner et al., 2022) and that this decrease in interpretation bias might be a mechanism of therapeutic change in CBT for anxiety (Mobach et al., 2021; Pereira et al., 2018; Steinman & Teachman, 2014; Waters et al., 2008) and mediate long-term outcomes of anxiety interventions (Makover et al., 2020).

Above, we focused on how poor safety learning relates to treatment outcome. Additionally, we argued that the tendency to *overgeneralize* based on prior learning experiences might constitute a vulnerability factor for developing anxiety (Lenaert et al., 2014). We illustrated this with the example of having a learning history with a volatile father, which could give rise to continuingly being on guard despite safe experiences and in the presence of, for instance, other men. In line with this view, promoting the discrimination between the CS+ (e.g., volatile father) and generalization stimuli (e.g., other men) could be a promising target in anxiety interventions. In other words, in a discrimination training individuals could learn that other men are different from the volatile father. Lommen et al. (2017) demonstrated in a fear conditioning study that such discrimination training could effectively reduce avoidance towards generalization stimuli. Similarly, in another fear conditioning study, promoting perceptual discrimination has found to result in decreased generalization of fear responding (Ginat-Frolich et al., 2017). In a first attempt to test the effects of discrimination training in a high-anxious sample (fear of spiders), Ginat-Frolich et al. (2019) demonstrated that the group receiving discrimination training showed less avoidance towards spiders with increased similarity to a live spider.

1.7.4 Conclusion

This chapter discussed how fear conditioning biases are related to pathological anxiety by providing a summary of patient-control differences in fear responding during conditioning procedures. In addition, we linked these biases in fear conditioning (as observable behavior) to interpretation biases (as a latent mental process). For example, in certain fear conditioning procedures (i.e., ‘weak’ situations) such as generalization tests and extinction trials, individuals vulnerable for anxiety might have a tendency to interpret ambiguous stimuli as threatening and as a consequence (still) respond fearfully to these (in fact safe) stimuli. Further empirical testing of whether and how interpretational processes impact fear conditioning is, however, needed to verify these predictions.

In providing different perspectives on biases, we explicitly distinguished biases as observable behavior (i.e., measured with experimental tasks) and biases as an underlying cognitive processes that could (partially) drive or mediate this observable behavior (also see De Houwer, 2019). We discussed the link between fear conditioning and interpretation biases in the light of a mediation model in which interpretation biases are an inborn and learned cognitive marker and processing style that could (partially) drive the observable differences in responding in fear conditioning procedures between individuals vulnerable for anxiety and controls.

Lastly, we discussed how these biases might impact the treatment of anxiety. In particular, we suggest that both biases might negatively impact treatment outcome. However, current evidence for this claim is mixed. Building further on this, techniques to remediate biases might enhance the efficacy of anxiety-focused interventions.

References

 Ahrens, L. M., Pauli, P., Reif, A., Mühlberger, A., Langs, G., Aalderink, T., & Wieser, M. J. (2016). Fear conditioning and stimulus generalization in patients with social anxiety disorder. Journal of Anxiety Disorders, 44, 36–46. [https://doi.org/10.1016/j.janxdis.2016.10.003](https://psycnet.apa.org/doi/10.1016/j.janxdis.2016.10.003)

Arnaudova, I., Krypotos, A. M., Effting, M., Boddez, Y., Kindt, M., & Beckers, T. (2013). Individual differences in discriminatory fear learning under conditions of ambiguity: A vulnerability factor for anxiety disorders? *Frontiers in Psychology*, *4*, 298. https://doi.org/10.3389/fpsyg.2013.00298

Arnaudova, I., Krypotos, A.-M., Effting, M., Kindt, M., & Beckers, T. (2017). Fearing shades of grey: Individual differences in fear responding towards generalisation stimuli. Cognition and Emotion, 31(6), 1181-1196. [https://doi.org/10.1080/02699931.2016.1204990](https://psycnet.apa.org/doi/10.1080/02699931.2016.1204990)

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Barrett, J., & Armony, J. L. (2009). Influence of trait anxiety on brain activity during the acquisition and extinction of aversive conditioning. *Psychological Medicine*, *39*(2), 255–265. https://doi.org/10.1017/S0033291708003516

Baumgardner, M., Silk, J.S., & Allen, K. B. (2022). Interpretation bias and anticipated distress in the face of ambiguity: Predictors of change in cognitive behavioral therapy for youth anxiety. *Child Psychiatry Hum Dev,* *53***,**479–488. https://doi.org/10.1007/s10578-021-01147-0

Beck, A. T., & Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behaviour Research and Therapy*, *35*(1), 49–58. https://doi.org/10.1016/s0005-7967(96)00069-1

Beckers, T., Krypotos, A. M., Boddez, Y., Effting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biological Psychology*, *92*(1), 90–96. https://doi.org/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. Journal of Behavior Therapy and Experimental Psychiatry, 44(2), 201–206. [https://doi.org/10.1016/j.jbtep.2012.08.003](https://psycnet.apa.org/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. *Cognition & Emotion*, *31*(7), 1444–1452. https://doi.org/10.1080/02699931.2016.1230541

Boddez, Y., Moors, A., Mertens, G., & De Houwer, J. (2020). Tackling fear: Beyond associative memory activation as the only determinant of fear responding. Neuroscience and Biobehavioral Reviews, 112, 410–419. [https://doi.org/10.1016/j.neubiorev.2020.02.009](https://psycnet.apa.org/doi/10.1016/j.neubiorev.2020.02.009)

Bouton, M. E. (1988). Context and ambiguity in the extinction of emotional learning: Implications for exposure therapy. *Behaviour Research and Therapy*, *26*(2), 137–149. https://doi.org/10.1016/0005-7967(88)90113-1

Bouton, M. E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*(10), 976–986. https://doi.org/10.1016/s0006-3223(02)01546-9

Britton, J. C., Lissek, S., Grillon, C., Norcross, M. A., & Pine, D. S. (2011). Development of anxiety: the role of threat appraisal and fear learning. *Depression and Anxiety*, *28*(1), 5–17. https://doi.org/10.1002/da.20733

Butler, E., Mobini, S., Rapee, R. M., Mackintosh, B., & Reynolds, S. A. (2015). Enhanced effects of combined cognitive bias modification and computerised cognitive behaviour therapy on social anxiety. Cogent Psychology, 2(1), Article 1011905. [https://doi.org/10.1080/23311908.2015.1011905](https://psycnet.apa.org/doi/10.1080/23311908.2015.1011905)

Cisler, J. M., Bacon, A. K., & Williams, N. L. (2009). Phenomenological characteristics of attentional biases towards threat: A critical review. *Cognitive Therapy and Research*, *33*(2), 221–234. https://doi.org/10.1007/s10608-007-9161-y

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, *30*(2), 203–216. https://doi.org/10.1016/j.cpr.2009.11.003

Cooper, S. E., Grillon, C., & Lissek, S. (2018). Impaired discriminative fear conditioning during later training trials differentiates generalized anxiety disorder, but not panic disorder, from healthy control participants. *Comprehensive Psychiatry*, *85*, 84–93. https://doi.org/10.1016/j.comppsych.2018.07.001

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy*, *46*(1), 5–27. https://doi.org/10.1016/j.brat.2007.10.003

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, *58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Craske, M. G., Treanor, M., Zbozinek, T. D., & Vervliet, B. (2022). Optimizing exposure therapy with an inhibitory retrieval approach and the OptEx Nexus. *Behaviour Research and Therapy*, *152*, 104069. https://doi.org/10.1016/j.brat.2022.104069

Cusack, K., Jonas, D. E., Forneris, C. A., Wines, C., Sonis, J., Middleton, J. C., Feltner, C., Brownley, K. A., Olmsted, K. R., Greenblatt, A., Weil, A., & Gaynes, B. N. (2016). Psychological treatments for adults with posttraumatic stress disorder: A systematic review and meta-analysis. Clinical Psychology Review, 43, 128–141. [https://doi.org/10.1016/j.cpr.2015.10.003](https://psycnet.apa.org/doi/10.1016/j.cpr.2015.10.003)

De Houwer J. D. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science*, *6*(2), 202–209. https://doi.org/10.1177/1745691611400238

De Houwer, J. (2019). Implicit bias is behavior: A functional-cognitive perspective on implicit bias. Perspectives on Psychological Science, 14(5), 835–840. [https://doi.org/10.1177/1745691619855638](https://psycnet.apa.org/doi/10.1177/1745691619855638)

de Jong, P. J., Merckelbach, H., & Arntz, A. (1995a). Covariation bias in phobic women: The relationship between a priori expectancy, on-line expectancy, autonomic responding, and a posteriori contingency judgment. Journal of Abnormal Psychology, 104(1), 55–62. [https://doi.org/10.1037/0021-843X.104.1.55](https://psycnet.apa.org/doi/10.1037/0021-843X.104.1.55)

de Jong, P. J., Merckelbach, H., Arntz, A., & Nijmam, H. (1992). Covariation detection in treated and untreated spider phobics. Journal of Abnormal Psychology, 101(4), 724–727. [https://doi.org/10.1037/0021-843X.101.4.724](https://psycnet.apa.org/doi/10.1037/0021-843X.101.4.724)

de Jong, P. J., Merckelbach, H., Bögels, S., & Kindt, M. (1998). Illusory correlation and social anxiety. Behaviour Research and Therapy, 36(11), 1063–1073. [https://doi.org/10.1016/S0005-7967(98)00099-0](https://psycnet.apa.org/doi/10.1016/S0005-7967%2898%2900099-0)

de Jong, P. J., van den Hout, M. A., & Merckelbach, H. (1995b). Covariation bias and the return of fear. *Behaviour Research and Therapy*, *33*(2), 211–213. https://doi.org/10.1016/0005-7967(94)e0024-d

Deacon, B., Kemp, J. J., Dixon, L. J., Sy, J. T., Farrell, N. R., & Zhang, A. R. (2013). Maximizing the efficacy of interoceptive exposure by optimizing inhibitory learning: A randomized controlled trial. *Behaviour Research and Therapy*, *51*(9), 588–596. https://doi.org/10.1016/j.brat.2013.06.006

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., van den Hout, M. A., & Baas, J. M. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety*, *32*(4), 239–253. https://doi.org/10.1002/da.22353

Duits, P., Richter, J., Baas, J. M. P., Engelhard, I. M., Limberg-Thiesen, A., Heitland, I., Hamm, A. O., & Cath, D. C. (2017). Enhancing effects of contingency instructions on fear acquisition and extinction in anxiety disorders. Journal of Abnormal Psychology, 126(4), 378–391. [https://doi.org/10.1037/abn0000266](https://psycnet.apa.org/doi/10.1037/abn0000266)

Dvir, M., Horovitz, O., Aderka, I. M., & Shechner, T. (2019). Fear conditioning and extinction in anxious and non-anxious youth: A meta-analysis. *Behaviour Research and Therapy*, *120*, 103431. https://doi.org/10.1016/j.brat.2019.103431

El-Bar, N., Laufer, O., Yoran-Hegesh, R., & Paz, R. (2017). Over-generalization in youth with anxiety disorders. Social Neuroscience, 12(1), 76–85. [https://doi.org/10.1080/17470919.2016.1167123](https://psycnet.apa.org/doi/10.1080/17470919.2016.1167123)

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

Garpenstrand, H., Annas, P., Ekblom, J., Oreland, L., & Fredrikson, M. (2001). Human fear conditioning is related to dopaminergic and serotonergic biological markers. Behavioral Neuroscience, 115(2), 358–364. [https://doi.org/10.1037/0735-7044.115.2.358](https://psycnet.apa.org/doi/10.1037/0735-7044.115.2.358)

Gazendam, F. J., Kamphuis, J. H., & Kindt, M. (2013). Deficient safety learning characterizes high trait anxious individuals. *Biological Psychology, 92*(2), 342–352. [https://doi.org/10.1016/j.biopsycho.2012.11.006](https://psycnet.apa.org/doi/10.1016/j.biopsycho.2012.11.006)

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

Ginat-Frolich, R., Klein, Z., Aderka, I. M., & Shechner, T. (2019). Reducing avoidance in adults with high spider fear using perceptual discrimination training. *Depression and Anxiety*, *36*(9), 859–865. https://doi.org/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. Behaviour Research and Therapy, 93, 29–37. [https://doi.org/10.1016/j.brat.2017.03.010](https://psycnet.apa.org/doi/10.1016/j.brat.2017.03.010)

Guthrie, R. M., & Bryant, R. A. (2006). Extinction learning before trauma and subsequent posttraumatic stress. *Psychosomatic Medicine*, *68*(2), 307–311. https://doi.org/10.1097/01.psy.0000208629.67653.cc

Haddad, A. D. M., 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. [https://doi.org/10.1007/s10862-012-9284-7](https://psycnet.apa.org/doi/10.1007/s10862-012-9284-7)

Haselton, M. G., Nettle, D., & Andrews, P. W. (2005). The evolution of cognitive bias. In D. M. Buss (Ed.), The handbook of evolutionary psychology (pp. 724–746). John Wiley & Sons, Inc..

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biological Psychiatry*, *60*(4), 361–368. https://doi.org/10.1016/j.biopsych.2005.10.006

Hirsch, C. R., Meeten, F., Krahé, C., & Reeder, C. (2016). Resolving ambiguity in emotional disorders: The nature and role of interpretation biases. *Annual Review of Clinical Psychology*, *12*, 281–305. https://doi.org/10.1146/annurev-clinpsy-021815-093436

Jinyao, Y., Xiongzhao, Z., Auerbach, R. P., Gardiner, C. K., Lin, C., Yuping, W., & Shuqiao, Y. (2012). Insecure attachment as a predictor of depressive and anxious symptomology. *Depression and Anxiety*, *29*(9), 789–796. https://doi.org/10.1002/da.21953

Kaczkurkin, A. N., Burton, P. C., Chazin, S. M., Manbeck, A. B., Espensen-Sturges, T., Cooper, S. E., Sponheim, S. R., & Lissek, S. (2017). Neural substrates of overgeneralized conditioned fear in PTSD. *The American Journal of Psychiatry*, *174*(2), 125–134. https://doi.org/10.1176/appi.ajp.2016.15121549

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62*(6), 593–602. [https://doi.org/10.1001/archpsyc.62.6.593](https://psycnet.apa.org/doi/10.1001/archpsyc.62.6.593)

Lange, I., Goossens, L., Michielse, S., Bakker, J., Vervliet, B., Marcelis, M., Wichers, M., van Os, J., van Amelsvoort, T., & Schruers, K. (2020). Neural responses during extinction learning predict exposure therapy outcome in phobia: Results from a randomized-controlled trial. *Neuropsychopharmacology*, *45*(3), 534–541. https://doi.org/10.1038/s41386-019-0467-8

Lau, J. Y., & Waters, A. M. (2017). Annual research review: An expanded account of information-processing mechanisms in risk for child and adolescent anxiety and depression. *Journal of Child Psychology and Psychiatry,* *58*(4), 387–407. https://doi.org/10.1111/jcpp.12653

Lenaert, B., Boddez, Y., Griffith, J. W., Vervliet, B., Schruers, K., & Hermans, D. (2014). Aversive learning and generalization predict subclinical levels of anxiety: A six-month longitudinal study. Journal of Anxiety Disorders, 28(8), 747–753. [https://doi.org/10.1016/j.janxdis.2014.09.006](https://psycnet.apa.org/doi/10.1016/j.janxdis.2014.09.006)

Lindert, J., von Ehrenstein, O. S., Grashow, R., Gal, G., Braehler, E., & Weisskopf, M. G. (2014). Sexual and physical abuse in childhood is associated with depression and anxiety over the life course: Systematic review and meta-analysis. *International Journal of Public Health*, *59*(2), 359–372. https://doi.org/10.1007/s00038-013-0519-5

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. *Biological Psychiatry*, *75*(11), 909–915. https://doi.org/10.1016/j.biopsych.2013.07.025

Lissek, S., Pine, D. S., & Grillon, C. (2006). The strong situation: A potential impediment to studying the psychobiology and pharmacology of anxiety disorders. Biological Psychology, 72(3), 265–270. [https://doi.org/10.1016/j.biopsycho.2005.11.004](https://psycnet.apa.org/doi/10.1016/j.biopsycho.2005.11.004)

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. *The American Journal of Psychiatry*, *167*(1), 47–55. https://doi.org/10.1176/appi.ajp.2009.09030410

Lommen, M., & Boddez, Y. (2022). Extinction learning as pretrauma vulnerability factor of posttraumatic stress: A replication study. *European Journal of Psychotraumatology*, *13*(1), 2051334. https://doi.org/10.1080/20008198.2022.2051334

Lommen, M., 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. https://doi.org/10.1371/journal.pone.0184485

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

Lommen, M. J. J., Engelhard, I. M., & van den Hout, M. A. (2010). Neuroticism and avoidance of ambiguous stimuli: Better safe than sorry? Personality and Individual Differences, 49(8), 1001–1006. [https://doi.org/10.1016/j.paid.2010.08.012](https://psycnet.apa.org/doi/10.1016/j.paid.2010.08.012)

Lonsdorf, T. B., & Kalisch, R. (2011). A review on experimental and clinical genetic associations studies on fear conditioning, extinction and cognitive-behavioral treatment. *Translational Psychiatry*, *1*(9), e41. https://doi.org/10.1038/tp.2011.36

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., . . . Merz, C. J. (2017). Don’t fear ‘fear conditioning’: Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. Neuroscience and Biobehavioral Reviews, 77, 247–285. [https://doi.org/10.1016/j.neubiorev.2017.02.026](https://psycnet.apa.org/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. https://doi.org/10.1016/j.neubiorev.2017.07.007

Lonsdorf, T. B., Weike, A. I., Nikamo, P., Schalling, M., Hamm, A. O., & Ohman, A. (2009). Genetic gating of human fear learning and extinction: possible implications for gene-environment interaction in anxiety disorder. *Psychological Science*, *20*(2), 198–206. https://doi.org/10.1111/j.1467-9280.2009.02280.x

Makover, H. B., Kendall, P. C., Olino, T., Carper, M. M., Albano, A. M., Piacentini, J., Peris, T., Langley, A. K., Gonzalez, A., Ginsburg, G. S., Compton, S., Birmaher, B., Sakolsky, D., Keeton, C., & Walkup, J. (2020). Mediators of youth anxiety outcomes 3 to 12 years after treatment. *Journal of Anxiety Disorders*, *70*, 102188. https://doi.org/10.1016/j.janxdis.2020.102188

Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A. E., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry*, *1*(5), 368–376. https://doi.org/10.1016/S2215-0366(14)70329-3

McLaughlin, N. C., Strong, D., Abrantes, A., Garnaat, S., Cerny, A., O'Connell, C., Fadok, R., Spofford, C., Rasmussen, S. A., Milad, M. R., & Greenberg, B. D. (2015). Extinction retention and fear renewal in a lifetime obsessive-compulsive disorder sample. *Behavioural Brain Research*, *280*, 72–77. https://doi.org/10.1016/j.bbr.2014.11.011

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

Mobach, L., Rapee, R. M., & Klein, A. M. (2021). The role of distorted cognitions in mediating treatment outcome in children with social anxiety disorder: A preliminary study. Child Psychiatry and Human Development. Advance online publication. [https://doi.org/10.1007/s10578-021-01268-6](https://psycnet.apa.org/doi/10.1007/s10578-021-01268-6)

Morey, R. A., Haswell, C. C., Stjepanović, D., Brancu, M., Beckham, J. C., Calhoun, P. S., Dedert, E., Elbogen, E. B., Fairbank, J. A., Tupler, L. A., Van Voorhees, E. E., Wagner, H. R., Kimbrel, N. A., Kirby, A., Marx, C. E., Kilts, J. D., Moore, S. D., Naylor, J. C., Swinkels, C., … LaBar, K. S. (2020). Neural correlates of conceptual-level fear generalization in posttraumatic stress disorder. *Neuropsychopharmacology*, *45*(8), 1380–1389. https://doi.org/10.1038/s41386-020-0661-8

Orr, S. P., Lasko, N. B., Macklin, M. L., Pineles, S. L., Chang, Y., & Pitman, R. K. (2012). Predicting post-trauma stress symptoms from pre-trauma psychophysiologic reactivity, personality traits and measures of psychopathology. *Biology of Mood & Anxiety Disorders*, *2*(1). https://doi.org/10.1186/2045-5380-2-8

Pauli, P., Montoya, P., & Martz, G. E. (2001). On-line and a posteriori covariation estimates in panic-prone individuals: Effects of a high contingency of shocks following fear-irrelevant stimuli. *Cognitive Therapy and Research*, *25*(1), 23–36. https://doi.org/10.1023/A:1026470514475

Pauli, P., Wiedemann, G., & Montoya, P. (1998). Covariation bias in flight phobics. *Journal of Anxiety Disorders*, *12*(6), 555–565. https://doi.org/10.1016/S0887-6185(98)00033-4

Pereira, A. I., Muris, P., Roberto, M. S., Marques, T., Goes, R., & Barros, L. (2018). Examining the mechanisms of therapeutic change in a cognitive-behavioral intervention for anxious children: The role of interpretation bias, perceived control, and coping strategies. *Child Psychiatry & Human Development*, *49*, 73–85. https://doi.org/10.1007/s10578-017-0731-2

Pineles, S. L., Vogt, D. S., & Orr, S. P. (2009). Personality and fear responses during conditioning: Beyond extraversion. *Personality and Individual Differences*, *46*(1), 48–53. https://doi.org/10.1016/J.PAID.2008.09.003

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. *Neuroscience & Biobehavioral Reviews*, *88*, 117–140. https://doi.org/10.1016/J.NEUBIOREV.2018.03.015

Pöhlchen, D., Leuchs, L., Binder, F. P., Blaskovich, B., Nantawisarakul, T., Topalidis, P., Brückl, T. M., Norrholm, S. D., Jovanovic, T., Spoormaker, V. I., Binder, E. B., Czisch, M., Erhardt, A., Grandi, N. C., Ilic-Cocic, S., Lucae, S., Sämann, P., & Tontsch, A. (2020). No robust differences in fear conditioning between patients with fear-related disorders and healthy controls. *Behaviour Research and Therapy*, *129*, 103610. https://doi.org/10.1016/j.brat.2020.103610

Pompoli, A., Furukawa, T. A., Efthimiou, O., Imai, H., Tajika, A., & Salanti, G. (2018). Dismantling cognitive-behaviour therapy for panic disorder: A systematic review and component network meta-analysis. *Psychological Medicine, 48*(12), 1945-1953. https://doi.org/10.1017/S0033291717003919

Raeder, F., Merz, C. J., Margraf, J., & Zlomuzica, A. (2020). The association between fear extinction, the ability to accomplish exposure and exposure therapy outcome in specific phobia. *Scientific Reports,10*(1), 4288. https://doi.org/10.1038/s41598-020-61004-3

Rattel, J. A., Miedl, S. F., Liedlgruber, M., Blechert, J., Seidl, E., & Wilhelm, F. H. (2020). Sensation seeking and neuroticism in fear conditioning and extinction: The role of avoidance behaviour. *Behaviour Research and Therapy*, *135*, 103761. https://doi.org/10.1016/J.BRAT.2020.103761

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations on the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (ed.), *Classical conditioning II: Current research and theory* (pp. 64-99) . Appleton-Century-Crofts.

Salemink, E., & van den Hout, M. (2010). Validation of the "recognition task" used in the training of interpretation biases. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 140-144. https://doi.org/10.1016/j.jbtep.2009.11.006

Scaini, S., Belotti, R., Ogliari, A., & Battaglia, M. (2016). A comprehensive meta-analysis of cognitive-behavioral interventions for social anxiety disorder in children and adolescents. *Journal of Anxiety Disorders*, *42*, 105–112. https://doi.org/10.1016/J.JANXDIS.2016.05.008

Scherer, K. R. (2021). Evidence for the existence of emotion dispositions and the effects of appraisal bias. *Emotion, 21*(6), 1224–1238. https://doi.org/10.1037/emo0000861

Scheveneels, S., Boddez, Y., & Hermans, D. (2019). Learning mechanisms in fear and anxiety. In *The Cambridge handbook of anxiety and related disorders* (pp. 13–40). Cambridge University Press. https://doi.org/10.1017/9781108140416.002

Scheveneels, S., Boddez, Y., & Hermans, D. (2021). Predicting clinical outcomes via human fear conditioning: A narrative review. *Behaviour Research and Therapy*, *142*, 103870. https://doi.org/10.1016/J.BRAT.2021.103870

Schoth, D. E., & Liossi, C. (2017). A systematic review of experimental paradigms for exploring biased interpretation of ambiguous information with emotional and neutral associations. *Frontiers in Psychology, 8*, 171. https://doi.org/10.3389/fpsyg.2017.00171

Sehlmeyer, C., Dannlowski, U., Schöning, S., Kugel, H., Pyka, M., Pfleiderer, B., Zwitserlood, P., Schiffbauer, H., Heindel, W., Arolt, V., & Konrad, C. (2011). Neural correlates of trait anxiety in fear extinction. *Psychological Medicine*, *41*(4), 789–798. https://doi.org/10.1017/S0033291710001248

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. *Neuroscience & Biobehavioral Reviews*, *107*, 320–328. https://doi.org/10.1016/J.NEUBIOREV.2019.09.029

Sijbrandij, M., Engelhard, I. M., Lommen, M. J. J., Leer, A., & Baas, J. M. P. (2013). Impaired fear inhibition learning predicts the persistence of symptoms of posttraumatic stress disorder (PTSD). *Journal of Psychiatric Research*, *47*(12), 1991–1997. https://doi.org/10.1016/J.JPSYCHIRES.2013.09.008

Spix, M., Lommen, M. J. J., & Boddez, Y. (2021). Deleting “fear” from “fear extinction”: Estimating the individual extinction rate via non-aversive conditioning. *Behaviour Research and Therapy*, *142*, 103869. https://doi.org/10.1016/J.BRAT.2021.103869

Steinman, S. A., & Teachman, B. A. (2014). Reaching new heights: Comparing interpretation bias modification to exposure therapy for extreme height fear. *Journal of Consulting and Clinical Psychology, 83*(3), 404-417. https://doi.org/10.1037/a0036023

Tinoco-González, D., Fullana, M. A., Torrents-Rodas, D., Bonillo, A., Vervliet, B., Blasco, M. J., Farré, M., & Torrubia, R. (2015). Conditioned fear acquisition and generalization in generalized anxiety disorder. *Behavior Therapy*, *46*(5), 627–639. https://doi.org/10.1016/J.BETH.2014.12.004

Tomarken, A. J., Mineka, S., & Cook, M. (1989). Fear-relevant selective associations and covariation bias. *Journal of Abnormal Psychology*, *98*(4), 381–394. https://doi.org/10.1037//0021-843x.98.4.381

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

Tzschoppe, J., Nees, F., Banaschewski, T., Barker, G. J., Büchel, C., Conrod, P. J., Garavan, H., Heinz, A., Loth, E., Mann, K., Martinot, J. L., Smolka, M. N., Gallinat, J., Ströhle, A., Struve, M., Rietschel, M., Schumann, G., & Flor, H. (2014). Aversive learning in adolescents: Modulation by amygdala-prefrontal and amygdala-hippocampal connectivity and neuroticism. *Neuropsychopharmacology*, *39*(4), 875–884. https://doi.org/10.1038/npp.2013.287

Van den Bergh, O., Brosschot, J., Critchley, H., Thayer, J. F., & Ottaviani, C. (2021). Better safe than sorry: A common signature of general vulnerability for psychopathology. *Perspectives on Psychological Science*, *16*(2), 225–246. https://doi.org/10.1177/1745691620950690

Van Overveld, M., De Jong, P. J., & Peters, M. L. (2010). Disgust and fear-related UCS-expectancy bias in blood-fearful individuals. *Clinical Psychology and Psychotherapy*, *17*(2), 100–109. https://doi.org/10.1002/cpp.639

Vervliet, B., & Boddez, Y. (2020). Aversive stimulus pairings are an unnecessary and insufficient cause of pathological anxiety. *Biological Psychiatry, 87*(10), 870–871. https://doi.org/10.1016/j.biopsych.2020.03.006

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

Waters, A. M., Wharton, T. A., Zimmer-Gembeck, M. J., & Craske, M. G. (2008). Threat-based cognitive biases in anxious children: Comparison with non-anxious children before and after cognitive behavioural treatment. *Behaviour Research and Therapy*, *46*(3), 358–374. https://doi.org/10.1016/J.BRAT.2008.01.002

Weisman, J. S., & Rodebaugh, T. L. (2018). Exposure therapy augmentation: A review and extension of techniques informed by an inhibitory learning approach. *Clinical Psychology Review*, *59*, 41–51. https://doi.org/10.1016/J.CPR.2017.10.010

Wiedemann, G., Pauli, P., & Dengler, W. (2001). A priori expectancy bias in patients with panic disorder. *Journal of Anxiety Disorders*, *15*(5), 401–412. https://doi.org/10.1016/S0887-6185(01)00072-X

Wiemer, J., & Pauli, P. (2016). Fear-relevant illusory correlations in different fears and anxiety disorders: A review of the literature. *Journal of Anxiety Disorders*, *42*, 113–128. https://doi.org/10.1016/J.JANXDIS.2016.07.003

Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. *Clinical Psychology Review, 28*(6), 1021-1037. https://doi.org/10.1016/j.cpr.2008.02.007

Würtz, F., Zahler, L., Blackwell, S. E., Margraf, J., Bagheri, M., & Woud, M. L. (2022). Scrambled but valid? The scrambled sentences task as a measure of interpretation biases in psychopathology: A systematic review and meta-analysis. *Clinical Psychology Review*, *93*, 102133. https://doi.org/10.1016/J.CPR.2022.102133

Yap, M. B. H., Pilkington, P. D., Ryan, S. M., & Jorm, A. F. (2014). Parental factors associated with depression and anxiety in young people: A systematic review and meta-analysis. *Journal of Affective Disorders*, *156*, 8–23. https://doi.org/10.1016/J.JAD.2013.11.007

**Acknowledgments**

We would like to thank the editor, Marcella Woud, for her thought-provoking suggestions. Funding: SS is employed on KU Leuven C1 project C16/19/02. YB is employed on Methusalem Grant BOF16/MET\_V/002.

1. Notably, as we will discuss below, when defining biases, this is one possible view on interpretation biases, namely as a mediating latent (mental) process. Alternatively, fear conditioning biases can be seen as one specific instance of interpretation bias, with both biases then situated at the level of behavior (also see De Houwer, 2019). In this chapter we will elaborate on interpretation bias as a mediating mental process that (partially) drives behavior in fear conditioning procedures. [↑](#footnote-ref-1)
2. Notably, these differences in acquisition and extinction between anxiety and controls have not always been replicated (Pöhlchen et al., 2020). [↑](#footnote-ref-2)
3. Notably, in a conceptual replication study in a sample of firefighters, Lommen and Boddez (2022) failed to replicate these results. [↑](#footnote-ref-3)
4. Note that in discussing the link with interpretation bias, interpretation bias was defined already as a mental process. [↑](#footnote-ref-4)
5. Note that interpretation bias can be defined as (1) observable behavior in experimental tasks such as a lexical decision task or (2) a latent mental processing style. In the mediation model that we propose here, interpretation bias is seen as a latent mental process. [↑](#footnote-ref-5)