Negative expectancy biases in psychopathology

de Jong, Peter; Daniels, Judith K.

Published in:
Cognitive Biases in Health and Psychiatric Disorders

DOI:
10.1016/B978-0-12-816660-4.00004-0

Chapter 4

Negative expectancy biases in psychopathology

Peter J. de Jong, Judith K. Daniels
Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Groningen, The Netherlands

Introduction

In an everchanging environment, it is crucial to extract regularities based on past events to understand the present and to predict the future. For effective adaptation, it is important to identify predictors and/or causes of pertinent outcomes such as bodily harm, food deprivation, or particular mental states (e.g., happiness, sadness, and fear). Such knowledge about the covariation of critical events allows to anticipate future outcomes, thereby increasing the odds to obtain desired ones (i.e., rewards) and avoid the aversive ones (i.e., punishment).

Consistent with the major functional importance of (valid) expectations, the brain has been conceptualized as a "prediction machine" that constantly compares predictions based on mental representations of expected associations (so-called priors) with actual outcomes (e.g., Fernández, Pedreira, & Boccia, 2017). The specific predictions coded in the priors are derived from probabilistic models, which rest on previous experiences with the current situation (and are thus highly learning dependent), or on inferred expectations (e.g., triggered by preceding cues). Incoming information is contrasted with the predicted information to enable rapid detection of unexpected events. Sensory input that is not consistent with these prior representations is thought to immediately become very salient. Unexpected events (i.e., prediction errors) are therefore also more likely to be remembered than predictable events, indicating that the human brain is particularly well adapted to react to a rapidly changing world (Proulx, Sleegers, & Tritt, 2017) (Fig. 1).

Together, this seems to imply that erroneous expectations will be readily corrected, thereby immunizing people against the development of robust expectancy biases. So how can we then explain that persistent and invalidating negative expectancy biases are very common and represent a core feature of virtually all mental disorders? What is at the core of anxiety, depression, and other

Cognitive Biases in Health and Psychiatric Disorders: Neurophysiological Foundations
https://doi.org/10.1016/B978-0-12-816660-4.00004-0

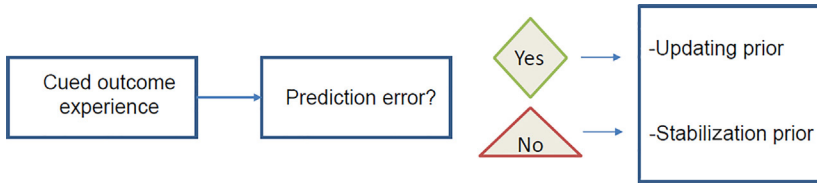

FIG. 1 Illustration of how outcome experiences may correct or reinforce generalized expectations (priors).

mental disorders that prevents proper updating of negative expectations? Which processes are involved in the apparent failure of learning from experience?

In the following section, we will first shortly describe the most common methods and measures that are used to investigate expectancy biases. Next, we will focus on the features of negative expectancy biases and how negative expectancies may emerge. Subsequently, we address in more depth why once established generalized negative expectancies are so resistant to change and how neural processes may be involved in this apparent insensitivity to disconfirming information.

Methods used to investigate expectancy biases

Studies that are designed to examine the role of expectancy biases in mental disorders typically rely on self-report measures to assess people's expectancies of particular events or outcomes. To investigate people's general inclination to expect particular (negative) outcomes, disorder-specific questionnaire measures have been developed. The items of such questionnaires typically take the form of conditional statements that specify the cue and/or context that may give rise to particular outcomes. For example, the Spider Phobia Beliefs Questionnaire (Arntz, Lavy, van den Berg, & van Rijsoort, 1993) asks people to indicate the probability of particular outcomes in situations involving the presence of a spider (e.g., during confrontation with a spider, the spider will bite me; if the spider does not go away and crawls on me, I will get a heart attack). Similarly, the Blushing Phobia Beliefs Questionnaire (Dijk, de Jong, Müller, & Boersma, 2010) asks people to indicate the probability of negative evaluations by others when they would display a blush (e.g., when I blush, others will think I am not socially skillful). As an alternative approach, some studies employed short vignettes describing disorder-relevant scenarios (e.g., an ambiguous interpersonal situation) and ask participants to indicate the probability of particular outcomes (e.g., I will be rejected; others will think I am incompetent) (e.g., Dijk & de Jong, 2012).

These questionnaire or vignette-based measures of people's expectations all rely on abstract or imagined stimuli or situations. As an alternative approach to investigate expectancy biases, lab-based methods have been designed in which people are exposed to concrete stimuli that are followed by concrete outcomes. Stimuli can, for example, represent disorder-relevant or

neutral (disorder-irrelevant) pictures or written scenarios, whereas aversive sounds, tactile (heat or electrical) stimuli, or rejecting faces are often used as negative outcomes (e.g., Duits et al., 2016; Hermann, Ofer, & Flor, 2004). During and after such procedure, participants can be asked to indicate their outcome expectancy given the presentation of a particular type of stimulus (e.g., de Jong, Merckelbach, & Arntz, 1995). Importantly, these lab-based procedures allow to experimentally manipulate the type of stimuli that are associated with aversive outcomes, the type and intensity of the aversive outcomes, and the objective stimulus/outcome contingencies. In other words, this type of measurement procedure allows to bring the factual experiences under experimental control, thereby setting the stage for investigating factors involved in the development or persistence of factually unjustified (dysfunctional) expectancies and for examining what factors may help correct/undermine biased expectancies. As an additional asset, these lab-based procedures allow to concurrently assess (neuro)physiological responses, which may help to further unravel the processes involved in the persistence of biased expectancies (e.g., Aue & Okon-Singer, 2015; Wiemer et al., 2015). The following sections will provide a series of concrete examples of how such procedures have been used in studies on negative expectancy biases.

Development of expectancy biases

Negative expectancy biases

There is ample evidence that unrealistic expectations are a core feature of psychopathology and may contribute to the development and chronicity of mental disorders (for a review, see Rief et al., 2015). These biased expectations can concern various aspects of future events (Fig. 2). First of all, expectancy biases often concern the probability that a particular negative outcome would

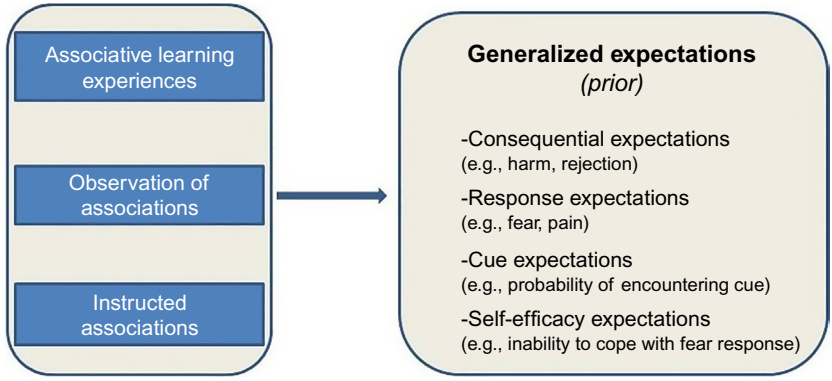

FIG. 2 Schematic illustration of the various pathways that may give rise to the development of various types of generalized expectations.

occur (e.g., when encountering a dog, it is very likely that the dog will bite me). However, expectancy bias may also concern the intensity/aversiveness of the anticipated negative outcome (e.g., such a dog bite will be extremely painful), and/or the costs that are associated with the occurrence of the anticipated negative outcome (e.g., I may get seriously wounded or may even die). In addition to these expectancies regarding specific consequences (*consequential expectations*), expectancy bias may concern the probability that a particular source of threat will occur in the first place (e.g., Aue & Hoeppli, 2012). In other words, people might overestimate the probability that one will encounter a threat cue such as a dog or will start blushing (i.e., *cue expectations*). Importantly, biased outcome expectancies not only are restricted to negative external events (such as a bite of a dog or signs of social rejection by others) but also may involve undesirable internal experiences such as pain and fear/panic (*response expectations* such as I will experience extreme fear/disgust/pain/despair) (see Peerdeman, van Laarhoven, Peters, & Evers, 2016). Finally, people might have biased expectations regarding their ability to cope with such internal or external threats (i.e., *self-efficacy expectations*).

The exact content and characteristics of biased expectations vary across disorders (Rief et al., 2015). For example, individuals with social anxiety disorder tend to overestimate the probability of displaying a blush (e.g., it is very likely that I will blush when I present my work) and are inclined to overestimate the negative consequences of their blushing (e.g., observers will think I am incompetent) (Dijk & de Jong, 2012). People with chronic pain disorder tend to expect that particular movements will damage their back (Vlaeyen, Crombez, & Linton, 2009); patients with Anorexia Nervosa expect that they will become fat (and therefore rejected) when they stop their excessive dieting (Levinson, Rapp, & Riley, 2014); people with spider phobia overestimate the probability of encountering a spider (Aue & Hoeppli, 2012; de Jong & Muris, 2002) and anticipate extreme fear and losing control when confronted with a spider (Arntz et al., 1993); patients with major depression typically exhibit negative expectations regarding their ability to deal with future stressful events (De Raedt & Hooley, 2016; for more disorder-specific examples, see Rief & Glombiewski, 2016).

A striking feature of expectancy biases is their tendency to persist even in the face of contradictory experiences. For example, patients with panic disorder typically hold on to the expectation that the next attack of palpitations will be fatal despite dozens of past panic attacks that turned out harmless. Clearly, such robustness of expectancy biases may help explain why mental disorders often run a chronic course. Yet, it also points to the crucial puzzle of why these expectations are so resistant to correction. To understand and effectively address the refractoriness of mental disorders, it is crucial to know why dysfunctional expectations tend to persist despite the availability of disconfirming evidence. To solve this conundrum, it may be helpful to first consider how negative expectations may generally arise.

Pathways to expectancy bias

Various pathways may foster individuals' (negative) expectancies (see Fig. 2). First, people may acquire expectancies via incidental learning experiences. On the basis of concrete experiences, people may learn that particular stimuli or behaviors are associated with particular (undesirable) consequences or outcomes. These associative learning experiences (e.g., being bitten by this particular dog) may give rise to more abstract generalized expectations or schemas (all dogs are dangerous and will bite me). These schemas are proposed to be formed by system consolidation, which is reflected in a gradual process of information reorganization and migration from hippocampus to neocortex (Fernández et al., 2017). Similar generalized expectations may also arise via observational (social) learning or modeling. For example, seeing one's mother panicking when confronted with a spider may result in a generalized expectation that spiders are dangerous animals. Third, negative expectancies may also be formed on the basis of instruction. For example, a parent may express concerns with regard to touching particular stimuli: "Don't pick up that cookie from the ground; it's dirty and will make you sick!," which may give rise to generalized contamination concerns.

To allow functional adaptation, emerging schemas should update their content and relations in the face of inconsistent (safe) experiences (*accommodation*). Importantly, the comparator (prediction error detection) function of our brain seems especially sensitive for detecting deviation in the more dangerous (aversive) direction and is relatively insensitive for detecting deviation below the expected level (Arntz, 1997). Accordingly, studies that experimentally manipulated participants' pain expectations showed that *under*predictions resulted in a very persistent heightened pain expectancy in the face of a series of disconfirming experiences (lower pain than expected), whereas such insensitivity to disconfirmation was absent in case of experimentally induced *over*predictions of pain. A single instance of higher pain than predicted was already sufficient to correct the apparently overoptimistic (i.e., underpredicted) pain expectations (e.g., Arntz, van Eck, & de Jong, 1992). Thus, the comparator system seems geared to a *better safe than sorry* heuristic.

Unexpected deviations in a more threatening direction may be taken as a red flag signaling that the next confrontation might exceed bearable limits and may be life threatening (Arntz, 1997). Such relatively high impact of underpredictions of threats can be seen as an adaptive mechanism since the implications of underpredictions seem far more critical for survival than overpredictions. Although this asymmetry with regard to the impact of over- vs underpredictions of threat may help avoid negative outcomes, it also implies that unexpected negative experiences have a disproportionally strong influence on the development of people's generalized expectancies and may thus set the stage for robust but objectively unjustified (i.e., biased) negative expectancies.

As soon as generalized negative expectations have been developed, several mechanisms may come into play that hamper correction of these expectations. In the following, we will subsequently highlight a series of mechanisms that have been put forward to explain the robustness of once acquired negative expectancy bias.

Factors that contribute to the robustness of expectancy bias

Avoidance behaviors

First of all, negative outcome or response expectations may motivate avoidance and other safety behaviors to ward off the anticipated threatening outcome. Convergently, several studies have shown that the strength of expectancy bias has predictive value for individuals' avoidance behaviors (e.g., Olatunji, Cisler, Meunier, Connolly, & Lohr, 2008). Avoidance behaviors hamper correction of unjustified generalized expectancies by preventing experiences that can disconfirm the validity of one's expectations. For example, although excessive hand washing in obsessive compulsive disorder (OCD) may be intended to avoid negative outcomes (e.g., the transmission of disease), it also serves to prevent the correction of inaccurate danger expectancies (e.g., touching a door knob is an important health risk). Accordingly, it has been shown that clinical interventions are less effective if patients are allowed to use this type of safety behaviors during exposure exercises (e.g., Sloan & Telch, 2002).

There is also experimental evidence within the context of aversive Pavlovian conditioning that points to avoidance as an important pathway to preserve once acquired negative expectations. For example, it has been shown that acquired shock expectancies (i.e., the CS^+ will be followed by a shock) persist if participants are subsequently given the opportunity to push a button upon the presentation of the CS^+ as a means to avoid the (expected) shock from being delivered. Using this button during the CS^+ only extinction procedure (avoidance) prevented participants to learn that in fact also without pressing the button, the shock was no longer delivered. Thus, avoidance interfered with updating (correction) of earlier acquired negative (shock) expectancies (e.g., Lovibond, Mitchell, Minard, Brady, & Menzies, 2009).

Importantly, negative outcome or response expectancies have been found to be strong predictors not only of avoidance but also of anticipatory anxiety (Rachman, 1990). For example, it has been shown that experimentally induced heightened pain expectations resulted in heightened anticipatory anxiety and physiological responding (as indexed by skin conductance and heart rate responses) (e.g., ; Arntz et al., 1992). Such heightened emotional responding in anticipation of a threatening outcome may not only contribute to avoidance motivation but also may be used as confirmatory information for the threat value of the anticipated outcome: If I feel anxious, there must be

danger (Arntz, Rauner, & van den Hout, 1995; Verwoerd, de Jong, Wessel, & van Hout, 2013). In other words, the anticipatory responses may be taken as further evidence for the validity of the biased outcome expectations, whereas subsequent avoidance will prevent the occurrence of disconfirming experiences that could have helped to correct the prior.

Subsequently, also the avoidance/escape behaviors themselves may be taken as evidence that indeed the occurrence of a negative outcome was prevented. In other words, employing safety behaviors not only may interfere with the modification of already existing expectations but also may promote the development of negative expectancies. Participants who were instructed to engage in OCD safety behaviors for 2 weeks showed an increase in danger estimates and fear of contamination (Deacon & Maack, 2008). One way to explain these findings is that these participants may have concluded that their safety behaviors might have prevented the occurrence of dangerous outcomes. This might have led to a threatening appraisal of initially harmless stimuli as dangerous. Thus, they might have inferred danger on the basis of their avoidance behaviors (if I feel the urge to wash, then there must be danger). In the context of nonexistent threats, such an "if I avoid, there must be danger," heuristic will logically contribute to the persistence of negative expectancy biases.

To test whether indeed patients with anxiety disorders infer danger on the basis of their own acts of avoidance, a group of participants with panic disorder, OCD, or social anxiety disorder and a nonclinical control group were asked to rate the danger in scenarios that systematically varied in the absence/presence of objective danger and the absence/presence of safety behavior (Gangemi, Mancini, & van den Hout, 2012). Interestingly, whereas nonanxious controls only relied on objective danger information, the danger estimates of anxiety patients were also influenced by safety behavior information. This tendency to infer danger on the basis of their avoidance behaviors might contribute to the development and persistence of phobic beliefs. As soon as one believes that one's safety behaviors imply danger, people may enter a downward spiral in which safety behaviors strengthen the perception of danger and vice versa.

All in all, there is consistent evidence that avoidance behaviors not only may contribute to the robustness of biased expectancies by preventing the experience of correcting events but also can contribute by functioning as input that strengthens prior (biased) expectations (see also Fig. 3).

Modulatory effects of expectancies

Negative expectancies may have an impact not only on strategically controlled behavior (e.g., avoidance) and reflective processes (e.g., emotional reasoning) but also on automatic, data-driven processes (e.g., saliency/perception threshold). Although people may generally conceive sensation and perception as passive bottom-up processes, as a mere registration of relevant internal or external stimuli, this is clearly not the case. There is ample evidence that our perception

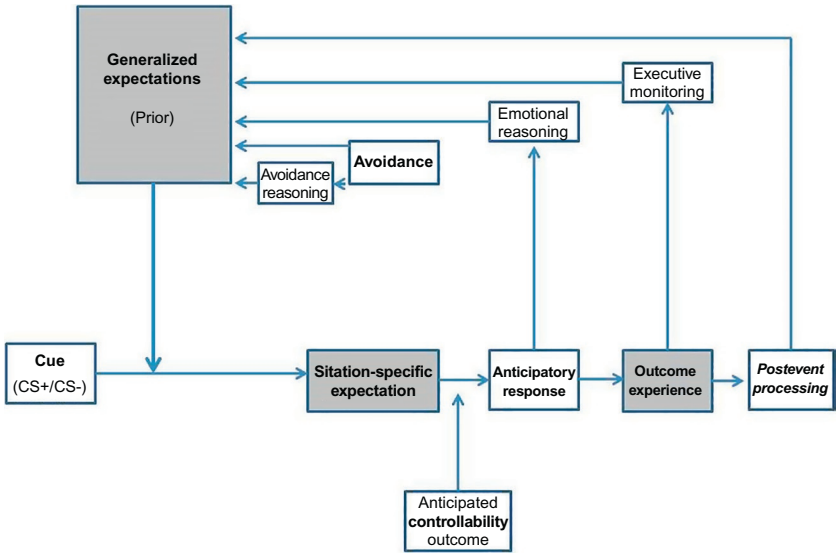

FIG. 3 Schematic illustration of how the proposed factors that are described in this section may contribute to robustness of (biased) generalized expectations. The concrete cue-based situation-specific expectations are moderated by the prior and may elicit avoidance and anticipatory responses. The type of anticipatory responses (e.g., fear: sensory preprocessing) will have an impact on the actual outcome experience. The type of impact will depend on the anticipated controllability of the outcome. Independent of the outcome, anticipatory emotional responses may give rise to emotional reasoning (e.g., if I feel anxious, it must be dangerous), which may confirm the prior. The concrete expectation may also directly motivate avoidance, thereby preventing the outcome to occur; because this precludes expectation-violating experiences, this may consolidate the prior, possibly also indirectly via avoidance-based reasoning bias (if I avoid, it must be dangerous). The concrete outcome experience is monitored (e.g., with regard to cue-outcome contingencies), which will either result in consolidation or modification of the prior. Finally, postevent processing of the actual experience will provide input to consolidate or modify the prior.

is influenced by our expectations, even already during very early processing steps. Moreover, there is increasing evidence indicating that also the process of *sensation* can be modulated by higher order cognitions such as expectations (e.g., Sussman, Mohanty, & Jin, 2020).

The concrete type of modulation depends on the functional meaning or goal relevance of the expected events. If, for example, it has functional value to readily detect the movement of a particular stimulus (e.g., a spider) when it is heading for a particular direction (e.g., toward oneself), expectancy-based sensory processing in the primary visual cortex can be a means to facilitate sensory processing of the moving stimulus (cf. Kok, Brouwer, van Gerven, & de Lange, 2013). Prior expectancies may also have an inhibitory influence on the processing of sensory information. For example, in the context of *placebo* analgesia research, it has been shown that when participants expect less pain due to an "analgesic" procedure (e.g., application of a creme), they also show less activity

in the primary somatosensory brain areas during cued anticipation of the pain stimulus (e.g., Elsenbruch et al., 2012). This makes sense from the perspective that the placebo-induced lowered pain expectation rendered the cued pain stimulus less important/less motivationally salient. Because the pain stimulus was expected to be reduced in pain intensity, it was no longer critical to invest restricted resources in processing the sensory information that was elicited by the cued pain stimulus.

If, on the other hand, participants expect more intense pain due to a particular procedure, such *nocebo*-induced heightened pain expectancy is associated with heightened activation of the salience network or pain matrix (insula, anterior cingulate cortex, primary, and secondary sensory cortex; Legrain, Iannetti, Plaghki, & Mouraux, 2011). In this case, the modulation of sensory processing is consistent with the goal to monitor whether the increased pain remains within acceptable limits. These differential modulatory effects of heightened versus lowered (pain) expectations help explain how the same sensory input may result in very different subjective experiences. This not only is relevant within the context of placebo/nocebo effects but also may more generally help explain how systematic differences in priors may give rise to differential perception and emotional processing of both the predictive cues and the actual aversive outcomes.

Uncontrollability of expected outcomes

At this stage, it seems important to differentiate between the impact of expectations that concern outcomes, which people tend to perceive as controllable, and the impact of expectations of outcomes, which might be perceived as less or not at all controllable (and thus accompanied by low self-efficacy expectations). Most studies that are designed to investigate the moderating influence of expectations on the (emotional) impact of aversive outcomes such as Pavlovian conditioning and placebo studies, rely on very specific and narrowly circumscribed aversive stimuli that are presented for a very short and predictable duration (e.g., 1-s aversive tactile electrical shock, 500-ms burst of 100 dB white noise, 3-s heat stimulation of particular preset intensity, and 1-s presentation of facial expression signaling rejection). Moreover, consistent with ethical constraints, participants are instructed that they can always stop their cooperation, which further strengthens the perceived controllability of the aversive outcome (e.g., if it becomes too intense, I just quit). Under these circumstances, the emotional impact of an aversive outcome is typically lower when it is expected (e.g., due to a signal such as a CS^+) than when it is presented unexpectedly.

Thus, if cued, participants' subjective and physiological responding to the aversive outcome (as indexed by fMRI, skin conductance, and heart rate) is typically reduced compared with noncued or invalidly cued aversive outcomes (e.g., Knight, Waters, King, & Bandettini, 2010). This expectancy-elicited reduction in responding is known as unconditioned response (UCR) diminution (Kimmel, 1966). In the

meantime, the response to the predicting cue (CS^+) is typically heightened and typically shows a negative relationship with the magnitude of the responses to the outcome (UCR) (Goodman, Harnett, & Knight, 2018). This heightened conditional response (CR) to the predicting cues is thought to reflect the effort associated with the preparatory processes that assist in coping with the aversive outcome. The UCR diminution is taken to reflect the reduced impact due to these coping mechanisms that could be employed because of the predictive quality of the aversive outcome (Goodman et al., 2018).

The finding that people's emotional responding toward cued (and thus expected) aversive outcomes is reduced is not restricted to inherently negative stimuli. Also when spider fearful participants were presented with validly cued pictures of spiders, they responded with less intense emotional responses than when presented with uncued spider stimuli or when spiders were presented when they expected to see neutral stimuli on the screen (Sebastiani, D'Alessandro, & Gemignani, 2014). Thus, when it concerns controllable threats/aversive events, heightened outcome expectations result in lowered emotional responding to the cued aversive events, probably because individuals are then in a position to effectively prepare for the upcoming aversive outcome. The resultant lowered subjective and physiological responding to the aversive outcome may in fact signal a reduction of associated costs: Heightened expectations of its occurrence reduce the aversive properties of the outcome and heighten the perceived self-efficacy to cope with this type of aversive outcomes. Thus, in the context of controllable threats, heightened outcome expectancies tend to result in lowered emotional responding, which in turn may correct inflated expectancies about the associated costs.

Although the impact of controllable/narrowly circumscribed aversive outcomes is generally reduced when expected, the opposite seems true for outcomes that involve some uncertainty about the outcome (e.g., with regard to the exact intensity, exact timing, or duration) and/or that are perceived as relatively threatening or uncontrollable (Goodman, Harnett, & Knight, 2018; Goodman, Harnett, Wheelock, et al., 2018). For example, it has been shown that a cued pain stimulus of high intensity resulted in persistently heightened subjective pain and physiological responding to subsequently presented pain stimuli of medium intensity if it was suddenly presented following a series of cued pain stimuli of medium intensity (Arntz et al., 1992). In addition, the sudden increase in pain intensity resulted in heightened (defensive) heart rate responsivity to subsequently presented cues. Thus, it seems that the induced uncertainty about the exact nature of the outcome resulted in heightened fear (as indexed by defensive responsivity to the cue) together with heightened attention for the pain stimulus, which in turn might have modulated the promulgation of painful somatosensory input throughout the sensory system, thereby enhancing neural gain and pain perception (cf. Almarzouki, Brown, Brown, Leung, & Jones, 2017; Fardo et al., 2017). More generally, there is evidence that cued prestimulus processing of predictable though (highly) feared outcomes may directly

modulate sensory processing in lower sensory cortices (Reynolds & Chelazzi, 2004). This enhancement of sensory processing of feared outcomes can be mediated by direct projections from the higher cortex to the sensory cortices or to the sensory thalamus (Wimmer et al., 2015).

There is further evidence suggesting that the impact of cued expectations on the *regulation of emotional responses* to aversive outcomes might be moderated by (un)controllability. Wood et al. (2015) specifically designed a study to determine the independent and joint impact of predictability and controllability on the human brain response. To test the influence of predictability, half of the aversive outcomes were cued. To test the impact of controllability, half of the participants were provided with the opportunity to use a button to terminate the aversive stimulation. When predictable outcomes were uncontrollable, the neural activity of the ventromedial prefrontal cortex (vmPFC) and hippocampus was enhanced. When, on the other hand, predictable outcomes were controllable, neural activity of these areas was reduced. The vmPFC and hippocampus are proposed to play an important role in the emotion regulation process (e.g., Goodman, Harnett, & Knight, 2018; Goodman, Harnett, Wheelock, et al., 2018). The heightened neural activity may thus reflect increased efforts to cope with uncontrollable aversive events. Dysfunction of this neural circuitry may result in a failure to successfully regulate the emotional response to uncontrollable stressors. In turn, this may fuel negative outcome expectancies and promote avoidance of situations/behaviors that are linked to this type of uncontrollable aversive events. Thus, this type of expectancies may, for example, promote avoidance of places where "uncontrollable" dogs may appear or medical interventions that may give rise to uncontrollable aversive side effects. From such perspective, it would be interesting to see in future research whether individual differences in participants' brain responses to uncontrollable outcomes have prognostic value for people's more general ability to cope with future stressors.

All in all, the available evidence may be taken to indicate that high controllability of cued outcomes may help prevent the development and persistence of negative outcome expectancies (via reduced physiological and emotional responding), whereas low (perceived) control of cued outcomes may promote the development of negative expectancy biases (via heightened physiological and emotional responding).

Uncertainty and illusory correlations

In most studies discussed so far, there was a clear contingency between the predicting stimulus (cue/CS^+) and the aversive outcome (unconditional stimulus, US). In a typical Pavlovian (differential aversive) conditioning procedure, the CS^- is never and the CS^+ is always followed by the negative US. However, in real life, the covariation between particular cues and particular outcomes is typically less straightforward/more ambiguous. This can have major implications. First of all, cues that have limited predictive power leave room for uncertainty

about whether or not the outcome will occur and whether or not one can cope with this situation. Accordingly, people generally respond with larger stress/anticipatory fear responses to cues with lower predictive power (e.g., cues that signal 20% or 60% probability of US occurrence) than to cues that are always (100%) followed by the US (e.g., Hefner & Curtin, 2012); this seems especially the case for high anxious individuals. In turn, these heightened anticipatory physiological fear responses may give rise to emotional reasoning and promote escape and avoidance, thereby contributing to the persistence of heightened expectancy bias.

Second, there is ample evidence that the perceived covariation between stimuli/cues and outcomes varies as a function of current situational information and prior expectations (e.g., Alloy & Tabachnik, 1984). Thus, when the current situational information is ambiguous, the assessment of covariations is typically determined by people's expectations; the stronger the expectations, the more are people inclined to perceive covariations that are consistent with their prior expectations, which in turn reinforces their prior expectations, etc.

In line with this, it has been shown that once expectations about covariations between a particular cue and the aversive outcome are induced (stimulus X is more often followed by shock than stimulus Y), these expectations are highly robust against refutation by new information that is factually inconsistent with this earlier acquired expectation (de Jong, Merckelbach, & Arntz, 1990). More specifically, it was shown that when two categories of cue stimuli were followed by an aversive outcome equally often (50%), the differential shock expectancies persisted and tended to become even stronger over trials. Following the experiment, participants generally overestimated the probability of shock given the target slide. This illusory correlation was also expressed in lowered skin conductance responses to the shock outcome on trials where the cue slides preceded the shock (UCR diminution), and relatively strong "omission" responses following the target slide when the cue slide was not followed by a shock outcome. In other words, participants seemed to display an orienting response upon the nonoccurrence of the expected outcome. Importantly, illusory correlations between cues and aversive outcomes have been shown to be most easily elicited when the outcome was relatively aversive (Wiemer, Mühlberger, & Pauli, 2014). Aversive outcomes might elicit a particularly strong motivation to detect covariations between cues and outcomes to help avoid future encounters. Also, aversive outcomes are typically the target of negative expectancies in patients with mental disorders. Thus, it seems reasonable to propose that the perception of illusory correlations may also contribute to the persistence of disorder-specific negative expectancy biases.

In line with this, there is consistent evidence that illusory correlations (IC) are involved in various mental disorders including spider phobia (e.g., de Jong, Merckelbach, Arntz, & Nijman, 1992), panic disorder (Pauli, Montoya, & Martz, 1996), social anxiety disorder (e.g., Hermann et al., 2004), posttraumatic stress disorder (Engelhard, de Jong, van den Hout, & van Overveld, 2009),

blood-injury fears (van Overveld, de Jong, & Peters, 2010), eating disorder (Mayer, Muris, Kramer-Freher, Stout, & Polak, 2012), and body dissatisfaction (Alleva, Martijn, & Jansen, 2016). In the typical IC study, participants are presented with a series of various types of stimuli (e.g., pictures of spiders, flowers, and mushrooms) that are followed by various outcomes (e.g., shock, tone, and nothing) (e.g., Tomarken, Sutton, & Mineka, 1995). Importantly, each stimulus-outcome combination is presented equally often. In spite of the absence of a systematic relationship between particular stimuli and particular outcomes, participants typically perceive a covariation between disorder-relevant stimuli and aversive outcomes. Disorder-specific ICs are reflected in both online US expectations and postexperimentally reported covariation estimates. Consistent with the view that these ICs may contribute to the persistence of mental disorders, ICs tend to reduce following successful treatment (e.g., van Overveld, de Jong, Huijding, & Peters, 2010), whereas residual ICs following treatment showed prognostic value for a return of fear (de Jong et al., 1995). As further indirect evidence for their causal influence, ICs showed predictive value for the persistence of PTSD symptomatology (Engelhard et al., 2009).

In apparent conflict with findings related to ICs between neutral stimuli and aversive outcomes, ICs between disorder-relevant stimuli (e.g., spiders) and aversive outcomes are paralleled with heightened instead of lowered UCRs to the aversive outcomes that are preceded by disorder-relevant "predictive" stimuli. Accordingly, it has been shown that in spite of the heightened shock expectancy on spider cue trials, skin conductance responses to shock outcomes were not reduced (no UCR diminution) but were in fact increased (e.g., de Jong et al., 1995; de Jong & Merckelbach, 1991). One explanation might be that under these circumstances, fearful individuals experience the aversive outcome as relatively uncontrollable. The heightened UCR may then reflect the increased effort to regulate the emotional response to uncontrollable stressors (see also the previous section). As a second explanation, it might be that the increased responding to the aversive outcome is due to affective response matching (vanOyen Witvliet & Vrana, 2000). In contrast to the typical UCR diminution studies, the predictive cues in the IC studies are not intrinsically neutral but elicit distress and defensive arousal also independent of their perceived predictive validity for shock outcome. Thus, the defensive response that is elicited by the disorder-relevant stimuli may in fact prime the defensive response to the aversive outcome stimulus (e.g., a shock or burst of white noise), resulting in a potentiated UCR.

Consistent with such view, it has been found that the IC between spider pictures and shock outcome was paralleled with relatively strong activity of the primary sensory motor cortex (PSMc) in response to aversive outcomes that were cued by spider pictures (Wiemer et al., 2015). In individuals with spider phobia, PSMc activity showed a positive correlation with the experienced aversiveness of the shock outcome. Thus, these findings may indicate that the spider cues primed the sensory processing of the subsequently presented shock resulting in

more intense processing and heightened aversiveness of the shock outcome. In its turn, the heightened aversiveness of the shock following spider pictures may contribute to the persistence of heightened shock expectancies on spider trials (cf. Wiemer et al., 2014).

Together, the available evidence indicates that also within a more ambiguous context of low or zero contingencies, persistent expectancy biases may arise. This seems especially the case when the US is relatively aversive and when the cue per se is already perceived as threatening; exactly, the type of features that is also typically involved in disorder-relevant expectations. Thus, it seems that also within a relatively ambiguous context, which is closer to the covariations in real life than modeled in a prototypical differential conditioning paradigm, robust expectancy biases may arise.

Postevent "validation" processes (rumination and immunization)

Finally, there are a series of processes that immunize against correction of expectancy biases, which take place in the aftermath of actual experiences. For example, people may start ruminating about earlier CS-US co-occurrences and/or about the aversiveness of the US. Experimental induction of such ruminative processes within the context of aversive conditioning procedures have been shown to result in both heightened fear of the predictive cue (CS^+) and relatively persistent US expectancies (e.g., Gazendam & Kindt, 2012; Joos, Vansteenwegen, Vervliet, & Hermans, 2013). It would be interesting to see in future research whether such experimental procedures would also result in stronger connectivity between the dorsolateral prefrontal cortex (dlPFC) and somatosensory cortex activity as a possible underlying neurophysiological process (cf. Greening, Lee, & Mather, 2016).

Postevent processes may also immunize against apparently straightforward refutations of prior expectations. For example, people may discard the invalidating evidence as an exceptional, nonrepresentative event and the exception that proofs the rule (Rief et al., 2015) or challenge the validity/relevance of the event for their expectation or put extremely heavy weight on past events to reduce the value of this single current event. This type of processes may thus result in an assimilation of the refuting evidence within the prior (the generalized expectancies), instead of accommodating the prior in response to the new (conflicting) evidence. To aggravate matters, people generally tend to rely on confirmation biased heuristics such as "what I believe is true" (e.g., Vroling, Glashouwer, Lange, Allart, & de Jong, 2016), which further hampers the correction of expectations on the basis of disconfirming evidence. In addition, people are generally inclined to selectively search for additional information that confirms prior expectations together with ignoring information that is at odds with their prior views (e.g., de Jong & Vroling, 2014). Such selective search for expectancy-confirming information will further hamper correction of distorted expectancies

and may thus contribute to the persistence of "pathogenic" expectancy biases (de Jong, 2015). Thus, within the context of treatment, it seems important to design strategies that help optimizing the impact of prediction errors and to minimize the opportunity for postevent immunization processes.

The neural basis of expectancy biases

In the previous section(s), we already alluded to brain processes that may be involved in the development and persistence of (biased) expectancies. This section zooms in on the neural basis of expectancy biases and aims to provide a global outline of the neural mechanisms that may be involved in the failure to correct generalized expectancies. Consistent with the conceptualization of the brain as a "prediction machine," which constantly compares predictions based on mental representations of expected associations with actual outcomes (e.g., Fernández et al., 2017), several brain regions have been implicated in these processes—ranging from regions subserving the low-level processing of sensory stimuli to evolutionary younger brain regions involved in executive functions.

Expectations influencing perception

In his book, *Principles of psychology*, James (1890) already suggested that perception is more than a direct registration of sensations and argued that although a part of what we perceive comes through our senses, another part (and it may be the larger part) always comes out of our head. Since then, many studies have shown that perception is far from a passive reflection of available sensory information. For example, recent neuroimaging studies indicate that the recognition of emotions and behavioral response times to targets in visual search paradigms might be influenced by prior expectations, likely due to a strategic deployment of attentional processes triggered by the preceding cues (Aue, Guex, Chauvigné, Okon-Singer, & Vuilleumier, 2019; Barbalat, Bazargani, & Blakemore, 2013; Dzafic, Martin, Hocking, Mowry, & Burianová, 2016; Sussman et al., 2020).

Moreover, there is now ample evidence indicating that even the process of *sensation* can be modulated by higher order cognitions (e.g., expectations). For example, if the brain expects sensory information stemming from the touch of the finger tips, the processing for this type of information is preactivated and thus sensitized (Fiorio & Haggard, 2005). This even works across sensory domains (facilitated by multimodal processing in the posterior parietal cortex) so that the visual presentation of an image of a hand enhances the subsequent sensory acuity for touch of the fingertips (Konen & Haggard, 2014). This enhancement functions not only at the central level but also even all the way down to the sensory organs. While interactions between sensory pathways such as the visual and auditory systems have long been known to occur in the sensory cortices (for a review Choi, Lee, & Lee, 2018), it was recently shown that prior knowledge of the directionality of a subsequent sound leads to adjustment of the eardrums

to optimize sound perception from this direction and that this adjustment is synchronized with the eye movements (Gruters et al., 2018). Together, this type of studies seems to indicate that integration of prior expectations and sensory input is already evident in the earliest stages of bottom-up sensory processing.

Neural correlates of overgeneralization of fear stimuli

Overgeneralization of fear stimuli is one of the mechanisms related to the development of biased expectations. The dlPFC has been associated with contingency awareness during fear conditioning (Carter, O'Doherty, Seymour, Koch, & Dolan, 2006), and a recent study indicates that an underactivation of the dlPFC could be associated with overgeneralization of fear cues in anxiety patients (Balderston, Hsiung, Ernst, & Grillon, 2017; cf. Wiemer et al., 2015). In addition, two recent studies indicated that overgeneralization of fear stimuli might be associated with an underactivation of the medial PFC during safe trials in both trauma exposed individuals (Harnett et al., 2018) and patients with generalized anxiety (Greenberg, Carlson, Cha, Hajcak, & Mujica-Parodi, 2013a), which might indicate that limited downregulation of limbic structures via prefrontal structures might be driving this process. Convergently, overgeneralization of fear has been repeatedly associated with an increased activation of the insula (Greenberg et al., 2013a; Greenberg, Carlson, Cha, Hajcak, & Mujica-Parodi, 2013b; Kaczkurkin et al., 2017; Morey et al., 2015; Tuominen et al., 2019).

Neural correlates of regulatory responses

Top-down regulation generally facilitates adaptive coping in stressful situations. However, there are large individual differences in how effective this top-down modulation is, and limitations in stress regulation are thought to play a central role in the development and clinical course of several mental disorders (Compas et al., 2017). When stressors are predictable, the adaptive regulatory response starts right after the cue and entails preparatory processes of up- or downregulation of several neurocircuits before the stressful event occurs. This preparatory response is associated with activations in the right dlPFC, which in turn is causally related to behavioral performance (Vanderhasselt et al., 2007). The elicited dlPFC activation also modulates the generation of autonomic (e.g., skin conductance) responses (see Remue et al., 2016, for an experimental TMS study), likely mediated via the ventromedial PFC (vmPFC), which has direct anatomical connections to the amygdala. As a result, nonreinforced trials that were cued as CS^+ trials are associated with an overactivation of the somatosensory cortex, which is functionally connected to the dlPFC at that moment (Greening et al., 2016). The right dlPFC thus plays a vital role both during the preparatory stage and contingency monitoring and the detection of conflict (i.e., prediction error).

The preparatory response is known to be strongly moderated by expectations. Under conditions of controllability, cued fear stimuli seem to elicit less

preparatory activation, particularly in the vmPFC and the hippocampus (Wood et al., 2015), which may reflect diminished efforts to prepare for the impending aversive event as the event can be avoided. If the anticipated aversive outcome cannot be avoided, the preparatory responses will probably be heightened as a reflection of the increased effort to cope with the upcoming threatening outcome (cf. Wiemer et al., 2015). However, whether indeed such increased preparatory response activation occurs may depend on people's self-efficacy expectations. It has been argued that if people have low expectations of successful coping, this could lead to an underdeployment of preparatory activation (De Raedt & Hooley, 2016). More generally, this seems to imply that expectations about one's options to deal with an upcoming stressful event may already shape the regulatory response before its onset.

Illusory correlations have been found to be associated with an overactivation of the left dlPFC directly during the presentation of phobia-specific cues (Wiemer et al., 2015, see Wiemer & Pauli, 2016b for results of a related functional connectivity analysis). In a group of phobia patients, the overactivation of the left dlPFC during the presentation of pictures with spiders predicted the illusory correlation between spider pictures and the likelihood of subsequent shocks. Interestingly, Aue et al. (2015) did not find a significant association between brain activity during the presentation of phobia-relevant cues and biased encounter expectancies. Analyzing brain activation during the subsequent rating phase (i.e., following stimulus processing) indicated that activation in the dlPFC correlated with the encounter expectancies: Phobic patients showed differential underactivation of the right dlPFC, the precuneus, and visual areas that were correlated with the strength of their encounter expectancy bias.

In conjunction, it seems clear that expectancy biases are associated with differential under- or overactivation of prefrontal regulatory brain structures. The exact patterns, however, still are somewhat inconsistent and need further exploration.

Somatovisceral responses related to negative expectancy biases

Evidence regarding autonomic responses that are associated with biased expectations comes mainly from lab studies using classical conditioning-like procedures. In the prototypical study, an originally neutral stimulus (CS^+) is paired with an aversive or painful outcome (e.g., electrical stimulation), whereas another type of stimulus (CS^-) is never paired with such an outcome. Most studies assessing autonomic responding in this type of research relied on participants' sweat response. The sweat response is under sympathetic control and can be easily assessed by putting constant voltage between two electrodes attached to participants' hand and to measure the current that runs between both electrodes (which will vary as a function of the enhanced conductivity of the hand when people show a sweat response) (for details, see Cacioppo, Tassinary, &

Berntson, 2017). The so-called skin conductance response (SCR) can be best seen as an index of general arousal, and depending on the context and type of stimulation, it may reflect a defensive reflex, an orientation response, cognitive effort, or preparatory efforts to cope with a certain upcoming situation. Clearly, then, it is not a "process-pure" measure and what exactly heightened SCRs reflect is not inherently evident from the response per se. Accordingly, heightened responding to the CS^+ may be interpreted not only as reflecting a defensive response but also as reflecting emotion regulation processes in anticipation of the upcoming aversive outcome (e.g., Grings, 1969).

Perhaps most relevant for the current context, it has been shown that people also show an SCR when following a series of reinforced CS^+ trials, the aversive outcome is incidentally omitted. Thus, at the offset of the CS^+, participants then show a so-called *omission response*, which may be interpreted as an orientation response elicited by the surprise that the aversive outcome did not occur (Seligman, Maier, & Solomon, 1971). Such an omission response may thus be used as an implicit measure of a negative outcome expectancy.

As already discussed in a previous section, also the SCR to the outcome itself may be affected by repeated pairings of a CS^+ and an aversive outcome. This so-called *UCR diminution* may reflect a weakened impact of predictable compared with unpredictable aversive outcomes (Goodman, Harnett, & Knight, 2018; Kimmel, 1966; Merckelbach & de Jong, 1988). Thus, also the SCR to the aversive outcome may be used as an implicit measure of negative outcome expectancies (the stronger the expectancy, the weaker the SCR to the outcome) (cf. Grings & Sukoneck, 1971). However, this UCR diminution may only be evident when it concerns a controllable/predictable outcome and when the signaling stimulus per se is affectively neutral. If the signal itself is considered as a threat (e.g., a spider) and the outcome as relatively uncontrollable (e.g., I would not know how to cope when the spider would really touch me), a cued aversive outcome would elicit a potentiated instead of an attenuated SCR (e.g., de Jong et al., 1995; vanOyen Witvliet & Vrana, 2000). Under these circumstances, it has been shown that relatively strong outcome expectations were accompanied by relatively large SCRs to the aversive event (e.g., de Jong et al., 1995). Thus, for using outcome elicited SCRs as an implicit measure of negative outcome expectations, it is crucial to take the type of cue (neutral or threatening) and type of outcome (controllable vs uncontrollable) into consideration.

As another way to index autonomic responding within the context of negative outcome expectancies, some studies included cardiovascular parameters with a focus on heart rate acceleration and deceleration. Because heart rate (HR) is under both sympathetic and parasympathetic control, HR acceleration can be due to an increase of sympathetic and/or a reduction of parasympathetic activity. Similarly, HR deceleration can be elicited by increased parasympathetic and/or reduced sympathetic activity. Most relevant to the current context, orientation responses are reflected in a HR deceleration, whereas defensive responses are reflected in a HR acceleration. Thus, if a neutral cue (CS^+) becomes

predictive of aversive outcomes, this will typically result in a CS^+-elicited HR deceleration, whereas the CS^-, which is never followed by an aversive outcome, will typically elicit weaker orientation responses (and thus a weaker HR deceleration) than the CS^+ (Arntz et al., 1992). However, if participants are uncertain about the intensity of the outcome (and thus with regard to their self-efficacy to cope with such outcomes), the CS^+ elicits a defensive attitude as is reflected in a HR acceleration instead of the typical HR deceleration that occurs in response to warning signals (CS^+) of outcomes that are of predictable intensity (e.g., Arntz et al., 1992). Thus, also for the cardiovascular measures it depends on the type of outcome (uncontrollable/uncertain intensity vs controllable/fixed intensity) whether expectancy biases may be evident in a relatively strong HR deceleration, or instead in a relatively strong HR acceleration in response to the CS^+. Finally, also for the cardiovascular measures there is evidence for UCR diminution. This is reflected in a weakened HR acceleration in response to the aversive outcome if the outcome was expected. However, also for HR responsivity yields that when the intensity of the outcome is uncertain and uncontrollable, cued outcomes elicit a HR acceleration (response potentiation) instead of a HR deceleration (response attenuation) (e.g., Arntz et al., 1992).

Similarities and differences between healthy and clinical populations

Negative expectancy biases represent common phenomena that are not restricted to people with mental disorders. For example, research using an illusory correlation paradigm showed that not only high but also low socially anxious individuals displayed persistently heightened expectancies that ambiguous social events (e.g., upon arrival on a party, an unknown person approaches you) or negative social events (e.g., at a party, you hear people gossiping about you) would be followed by social rejection (de Jong, de Graaf-Peters, van Hout, & van Wees, 2009). Similarly, both high and low spider fearful individuals indicated to expect that especially spider slides would be followed by an aversive shock outcome (e.g., de Jong, 1993). In the meantime, it seems that inflated cue expectancies (i.e., the probability that one would encounter a source of threat such as a spider) is restricted to fearful individuals (Aue & Okon-Singer, 2015), although it should be noted that research within this domain is thus far very limited.

Healthy and clinical populations are both subject to similar types of processes that counterforce modification of strong beliefs, especially if these beliefs concern the likelihood that particular signals predict harmful outcomes (e.g., de Jong, 2015). Such resistance to correct negative expectancies can be generally considered as an adaptive strategy because in case of life or death, it seems wise to play it safe. Moreover, in many cases, overly negative expectancies do not interfere with daily life. For example, for most people with negatively biased expectancies about what would happen upon encountering a spider, this bias

has no impact on their quality of life because they can easily avoid or escape situations where they may encounter a spider. Under such circumstances, there is little to gain by testing whether indeed one's negative expectancies are factual correct.

Obviously, the situation is fundamentally different for people with mental disorders. By definition, their outcome expectancies strongly interfere with their daily activities and have a highly negative impact on their quality of life. In these cases, the common insensitivity to threat-disconfirming information may render patients' immune for information that could help correct their invalidating expectancy biases. On the basis of the currently available evidence, it remains unclear whether the differences between clinical and nonclinical populations are due to differences in the strength of the expectancies (e.g., due to differences in learning history) or to fundamental differences in neurological or cognitive functioning that render patients especially prone to threat-confirming information processing. Irrespective of the ultimate source of the differences in expectancy biases between clinical and nonclinical populations, it is clear that patients with mental disorders fail to correct disorder-relevant expectancy biases in spite of their detrimental influence on their quality of life. Thus, from a clinical perspective, it is critical to design interventions that promote an "advocate of the devil" perspective and that help counterforce the various processes that could otherwise contribute to the robustness of patients' dysfunctional and factually unjustified negative expectancies.

Recapitulation, limitations, and some future directions

Negative expectancy biases can be conceived as inadvertent by-products of otherwise functional processes that help prevent the occurrence of negative (harmful) outcomes. Negative expectancy biases represent common phenomena, and are at the core of many mental disorders. To correct dysfunctional expectancy biases within the context of mental disorders, common treatment procedures such as exposure in vivo and behavioral experiments can be conceptualized as opportunities to elicit salient prediction errors. Starting point of such an approach is thus to elicit expectation-violating experiences, which in turn would lead to adjustment of the generalized expectations. Yet, disorder-specific expectancies appear highly robust against correction and even tend to be immune against straightforward refutations. Thus, to improve treatment efficacy, it would be important to identify factors that critically contribute to the robustness of patients' expectancies (cf. Kang, Vervliet, Engelhard, van Dis, & Hagenaars, 2018; Kube, Rief, Gollwitzer, Gärtner, & Glombiewski, 2018).

This chapter discussed a series of candidate mechanisms that may help explain why once acquired generalized negative expectations are so robust against refutation. These factors included avoidance of possible expectation-violating situations, unjustified inferences on the basis of anticipatory distress and avoidance behaviors (if I avoid something, it must be dangerous), expectancy-based

modulation of sensory input and perceptual processing, the perception of expectancy-congruent illusory correlations, and postevent "immunizing" processes such as rumination. The reviewed literature not only provided helpful starting points for explaining the refractoriness of expectancy biases but also pointed to some important gaps in the available evidence that await further scrutiny. In the following, we outline just a couple of leads that seem worth of further investigation.

Importantly, patients' expectations are not only robust but also overgeneral and sensitive for illusory correlations. In the previous section, several neural circuits have been proposed to be involved in the persistence of negative expectancy biases. Most prominently, it was discussed that regulatory regions in the dorsolateral PFC show altered activation depending on the cued expectations as well as disorder-specific aberrations. Since the dlPFC has been shown to be involved in contingency awareness, it is tempting to assume that altered prefrontal control could also counterforce the development of objectively unjustified cue-outcome associations.

In addition, it was discussed that cued expectations of aversive outcomes resulted in anticipatory somatosensory cortex activity that was associated with heightened aversiveness of the actual outcome (e.g., Wiemer et al., 2015). Interestingly, the strength of cued sensory preprocessing showed a positive correlation with trait anxiety (Greening et al., 2016). Individuals with high trait anxiety may thus experience cued aversive events as more aversive than people low in trait anxiety. This may help explain why trait anxiety is associated with a heightened chance of developing biased outcome expectancies and anxiety disorders (e.g., Chan & Lovibond, 1996). It should be emphasized, however, that most of the cited neuroimaging studies are purely correlational and cross-sectional and thus do not allow causal inferences. Therefore, an important and exciting next step for future research would be to bring those structures showing diverging activation patterns under experimental control, for example, via transcranial magnetic stimulation (TMS). In addition, it would be interesting and clinically relevant to test whether the strength of cue-based processing has prognostic value for treatment success and long-term prognosis (cf. Duits et al., 2016).

Finally, from the perspective that generalized expectancies would exert top-down influence on all levels of the system, it would also be relevant to see how exactly expectancy biases relate to other types of information processing biases that are often found in mental disorders such as attentional bias, interpretation bias, and memory bias. For understanding how to best modify robust expectancy bias, it would be especially relevant to know if these other types of biases represent relatively independent phenomena or can be better understood as functionally related biases that are in fact reflections of the modulating influence of biased expectancies. As a first step, it would be helpful to examine how the various biases relate and covary with regard to particular concerns (see, e.g., Aue & Okon-Singer, 2015).

As one example of such approach, a recent study was designed to test how expectancy bias and memory bias are related within the context of social anxiety (Caouette et al., 2015). Participants underwent a two-visit task that measured expectations about (visit 1) and memory of (visit 2) social feedback from unknown peers. In this study, it was found that the relationship between social anxiety and biased memory about the (negative) feedback was mediated by negative expectancy bias. Thus, these findings are consistent with the view that memory bias might be better conceptualized as a consequence of expectancy bias than as an independent agent. Further work along these lines including also attentional bias and interpretation bias and targeting also other types of concerns would be extremely welcome.

To more directly test to what extent heightened expectancies can indeed causally influence the strength of memory bias, it would be critical to experimentally vary the strength of participants' expectancies (e.g., by varying the objective contingencies) and to examine how this affects participants' memory bias. Recent work of Aue and colleagues exactly used such type of approach to test the impact of cued expectancies on attentional bias within the context of spider phobia (e.g., Aue, Chauvigné, Bristle, Okon-Singer, & Guex, 2016; Aue et al., 2019). Another approach could be to design a cognitive bias modification procedure that can specifically target expectancy bias and to examine the impact of such expectancy bias modification on other biases such as attentional, interpretation, and memory biases (cf. Everaert, Duyck, & Koster, 2014). Clearly, lot of future work is still needed to fully solve the critical puzzle of how exactly the various biases are interrelated and to what extent these relationships may vary across disorders.

To conclude, negative expectancy biases are assumed to play a critical role in the chronicity of mental disorders. Understanding what factors contribute to the robustness of expectancy biases is thus key for improving currently available treatment procedures. The available literature not only provided some important clues about relevant candidate mechanisms but also pointed to important questions that require further thought and further research. We hope that this chapter is helpful in inspiring future research to arrive at more final answers about the mechanisms involved in the persistence of negative expectancy bias and about the procedures that are most effective in adjusting invalidating negative expectancy biases in patients with mental disorders.

References

Alleva, J. M., Martijn, C., & Jansen, A. (2016). Covariation bias in women with a negative body evaluation: How is it expressed and can it be diminished? *Journal of Behavior Therapy and Expermental Psychiatry*, *50*, 33–39.

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, *91*, 112–149.

Almarzouki, A. F., Brown, C. A., Brown, R. J., Leung, M. H. K., & Jones, A. K. P. (2017). Negative expectations interfere with the analgesic effect of safety cues on pain perception by priming the cortical representation of pain in the midcingulate cortex. *PLoS One*, *12*, e0180006.

Arntz, A. (1997). The match-mismatch model of phobia acquisition. In G. C. L. Davey (Ed.), *Phobias: A handbook of theory, research and treatment* (pp. 375–395): Wiley.

Arntz, A., Lavy, E., van den Berg, G., & van Rijsoort, S. (1993). Negative beliefs of spider phobics: A psychometric evaluation of the Spider Phobia Beliefs Questionnaire. *Advances in Behavior Research and Therapy*, *15*, 257–277.

Arntz, A., Rauner, M., & van den Hout, M. (1995). "If I feel anxious, there must be danger": Ex-consequentia reasoning in inferring danger in anxiety disorders. *Behaviour Research and Therapy*, *33*, 917–925.

Arntz, A., van Eck, M., & de Jong, P. J. (1992). Unpredictable sudden increase of pain intensity and acquired fear. *Journal of Psychophysiology*, *6*, 54–64.

Aue, T., Chauvigné, L. A. S., Bristle, M., Okon-Singer, H., & Guex, R. (2016). Expectancy influences on attention to threat are only weak and transient: Behavioral and physiological evidence. *Biological Psychology*, *121*, 173–186.

Aue, T., Guex, R., Chauvigné, L. A. S., Okon-Singer, H., & Vuilleumier, P. (2019). Expectancies influence attention to neutral but not necessarily to threatening stimuli: An fMRI study. *Emotion*, *19*, 1244–1258.

Aue, T., & Hoeppli, M. E. (2012). Evidence for an encounter expectancy bias in fear of spiders. *Cognition and Emotion*, *26*, 727–736.

Aue, T., Hoeppli, M. E., Piguet, C., Hofstetter, C., Rieger, S. W., & Vuilleumier, P. (2015). Brain systems underlying encounter expectancy bias in spider phobia. *Cognitive and Affective Behavioral Neuroscience*, *15*, 335–348.

Aue, T., & Okon-Singer, H. (2015). Expectancy biases in fear and anxiety and their link to biases in attention. *Clinical Psychology Review*, *42*, 83–95.

Balderston, N. L., Hsiung, A., Ernst, M., & Grillon, C. (2017). Effect of threat on right dlPFC activity during behavioral pattern separation. *Journal of Neuroscience*, *37*, 9160–9171.

Barbalat, G., Bazargani, N., & Blakemore, S. J. (2013). The influence of prior expectations on emotional face perception in adolescence. *Cerebral Cortex*, *23*, 1542–1551.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2017). *Handbook of psychophysiology*. New York, NY: Cambridge University Press.

Caouette, J. D., Ruiz, S. K., Lee, C. C., Anbari, Z., Schriber, R. A., & Guyer, A. E. (2015). Expectancy bias mediates the link between social anxiety and memory bias for social evaluation. *Cognition and Emotion*, *29*, 945–953.

Carter, R. M., O'Doherty, J. P., Seymour, B., Koch, C., & Dolan, R. J. (2006). Contingency awareness in human aversive conditioning involves the middle frontal gyrus. *Neuroimage*, *29*, 1007–1012.

Chan, C. K., & Lovibond, P. F. (1996). Expectancy bias and trait anxiety. *Journal of Abnormal Psychology*, *105*, 637–647.

Choi, I., Lee, J.-Y., & Lee, S.-H. (2018). Bottom-up and top-down modulation of multisensory integration. *Current Opinion in Neurobiology*, *52*, 115–122.

Compas, B. E., Jaser, S. S., Bettis, A. H., Watson, K. H., Gruhn, M. A., Dunbar, J. P., et al. (2017). Coping, emotion regulation, and psychopathology in childhood and adolescence: A meta-analysis and narrative review. *Psychological Bulletin*, *143*, 939–991.

de Jong, P. J. (1993). Covariation bias in phobia: Mere resistance to preexperimental expectancies? *Behavior Therapy*, *24*, 447–454.

de Jong, P. J. (2015). Danger-confirming reasoning and the persistence of phobic beliefs. In N. Galbraith (Ed.), *Aberrant beliefs and reasoning* (pp. 132–153). London: Psychology Press.

de Jong, P. J., de Graaf-Peters, V., van Hout, W.J.P.J., & van Wees, R. (2009). Covariation bias for social events and signs of rejection in high and low socially anxious individuals. *Journal of Behavior Therapy and Experimental Psychiatry*, *40*, 359–373.

de Jong, P. J., & Merckelbach, H. (1991). Covariation bias and electrodermal responding before and after behavioural treatment. *Behaviour Research and Therapy*, *29*, 307–314.

de Jong, P. J., Merckelbach, H., & Arntz, A. (1990). Illusory correlation, on-line probability estimates, and electrodermal responding in a (quasi)-conditioning paradigm. *Biological Psychology*, *31*, 201–212.

de Jong, P. J., Merckelbach, H., & Arntz, A. (1995). 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*, 55–62.

de Jong, P. J., Merckelbach, H., Arntz, A., & Nijman, H. (1992). Covariation detection in treated and untreated spider phobics. *Journal of Abnormal Psychology*, *101*, 724–727.

de Jong, P. J., & Muris, P. (2002). Spider phobia interaction of disgust and perceived likelihood of involuntary physical contact. *Journal of Anxiety Disorders*, *16*, 51–65.

de Jong, P. J., & Vroling, M. S. (2014). Better safe than sorry: Threat-confirming reasoning bias in anxiety disorders. In I. Blanchette (Ed.), *Emotion and reasoning* (pp. 22–43). London: Psychology Press.

De Raedt, R., & Hooley, J. M. (2016). The role of expectancy and proactive control in stress regulation: A neurocognitive framework for regulation expectation. *Clinical Psychology Review*, *45*, 45–55.

Deacon, B. J., & Maack, D. J. (2008). The effects of safety behaviors on the fear of contamination: An experimental investigation. *Behaviour Research and Therapy*, *46*, 537–547.

Dijk, C., & de Jong, P. J. (2012). Blushing-fearful individuals overestimate the costs and probability of their blushing. *Behaviour Research and Therapy*, *50*, 158–162.

Dijk, C., de Jong, P. J., Müller, E., & Boersma, W. (2010). Blushing-fearful individuals' judgmental biases and conditional cognitions: An internet inquiry. *Journal of Psychopathology and Behavioral Assessment*, *32*, 264–270.

Duits, P., Klein Hofmeijer-Sevink, M., Engelhard, I. M., Baas, J. M. P., Ehrismann, W. A. M., & Cath, D. C. (2016). Threat expectancy bias and treatment outcome in patients with panic disorder and agoraphobia. *Journal of Behavior Therapy and Experimental Psychiatry*, *52*, 99–104.

Dzafic, I., Martin, A. K., Hocking, J., Mowry, B., & Burianová, H. (2016). Dynamic emotion perception and prior expectancy. *Neuropsychologia*, *86*, 131–140.

Elsenbruch, S., Kotsis, V., Benson, S., Rosenberger, C., Reidick, D., Schedlowski, M., et al. (2012). Neural mechanisms mediating the effects of expectation in visceral placebo analgesia: An fMRI study in healthy placebo responders and nonresponders. *Pain*, *153*, 382–390.

Engelhard, I. M., de Jong, P. J., van den Hout, M. A., & van Overveld, M. (2009). Expectancy bias and the persistence of posttraumatic stress. *Behaviour Research and Therapy*, *47*, 887–892.

Everaert, J., Duyck, W., & Koster, E. H. W. (2014). Attention, interpretation, and memory biases in subclinical depression: A proof-of-principle test of the combined cognitive biases hypothesis. *Emotion*, *14*, 331–340.

Fardo, F., Auksztulewicz, R., Allen, M., Dietz, M. J., Roepstorff, A., & Friston, K. J. (2017). Expectation violation and attention to pain jointly modulate neural gain in somatosensory cortex. *Neuroimage*, *153*, 109–121.

Fernández, R. S., Pedreira, M. E., & Boccia, M. M. (2017). Does reconsolidation occur in natural settings? Memory reconsolidation and anxiety disorders. *Clinical Psychology Review*, *57*, 45–58.

Fiorio, M., & Haggard, P. (2005). Viewing the body prepares the brain for touch: Effects of TMS over somatosensory cortex. *European Journal of Neuroscience*, *22*, 773–777.

Gangemi, A., Mancini, F., & van den Hout, M. (2012). Behavior as information: "If I avoid, then there must be a danger". *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, 1032–1038.

Gazendam, F. J., & Kindt, M. (2012). Worrying affects associative fear learning: A startle fear conditioning study. *PLoS One*, *7*, e34882.

Goodman, A. M., Harnett, N. G., & Knight, D. C. (2018). Pavlovian conditioned diminution of the neurobehavioral response to threat. *Neuroscience & Biobehavioral Reviews*, *84*, 218–224.

Goodman, A. M., Harnett, N. G., Wheelock, M. D., Hurst, D. R., Orem, T. R., Gossett, E. W., et al. (2018). Anticipatory prefrontal cortex activity underlies stress-induced changes in Pavlovian fear conditioning. *Neuroimage*, *174*, 237–247.

Greenberg, T., Carlson, J. M., Cha, J., Hajcak, G., & Mujica-Parodi, L. R. (2013a). Ventromedial prefrontal cortex reactivity is altered in generalized anxiety disorder during fear generalization. *Depression & Anxiety*, *30*, 242–250.

Greenberg, T., Carlson, J. M., Cha, J., Hajcak, G., & Mujica-Parodi, L. R. (2013b). Neural reactivity tracks fear generalization gradients. *Biological Psychology*, *92*, 2–8.

Greening, S. G., Lee, T. H., & Mather, M. (2016). Individual differences in anticipatory somatosensory cortex activity for shock is positively related with trait anxiety and multisensory integration. *Brain Sciences*, *6*, 2.

Grings, W. W. (1969). Anticipatory and preparatory electrodermal behavior in paired stimulation situations. *Psychophysiology*, *5*, 597–611.

Grings, W. W., & Sukoneck, H. I. (1971). Prediction probability as a determiner of anticipatory and preparatory electrodermal behavior. *Journal of Experimental Psychology*, *91*, 310–317.

Gruters, K. G., Murphy, D. L. K., Jenson, C. D., Smith, D. W., Shera, C. A., & Groh, J. M. (2018). The eardrums move when the eyes move: A multisensory effect on the mechanics of hearing. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E1309–E1318.

Harnett, N. G., Ference, E. W., Wood, K. H., Wheelock, M. D., Knight, A. J., & Knight, D. C. (2018). Trauma exposure acutely alters neural function during Pavlovian fear conditioning. *Cortex*, *109*, 1–13.

Hefner, K. R., & Curtin, J. J. (2012). Alcohol stress response dampening: Selective reduction of anxiety in the face of uncertain threat. *Journal of Psychopharmacology*, *26*, 232–244.

Hermann, C., Ofer, J., & Flor, H. (2004). Covariation bias for ambiguous social stimuli in generalized social phobia. *Journal of Abnormal Psychology*, *113*, 646–653.

James, W. (1890). *Principles of psychology*. New York: Holt.

Joos, E., Vansteenwegen, D., Vervliet, B., & Hermans, D. (2013). Repeated activation of a CS-US-contingency memory results in sustained conditioned responding. *Frontiers in Psychology*, *4*, 305.

Kaczkurkin, A. N., Burton, P. C., Chazin, S. M., Manbeck, A. B., Espensen-Sturges, T., Cooper, S. E., et al. (2017). Neural substrates of overgeneralized conditioned fear in PTSD. *American Journal of Psychiatry*, *174*, 125–134.

Kang, S., Vervliet, B., Engelhard, I. M., van Dis, E. A. M., & Hagenaars, M. A. (2018). Reduced return of threat expectancy after counterconditioning versus extinction. *Behaviour Research and Therapy*, *108*, 78–84.

Kimmel, H. D. (1966). Inhibition of the unconditional response in classical conditioning. *Psychological Review*, *73*, 232–240.

Knight, D. C., Waters, N. S., King, M. K., & Bandettini, P. A. (2010). Learning-related diminution of unconditioned SCR and fMRI signal responses. *Neuroimage*, *49*, 843–848.

Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, *33*, 16275–16284.

Konen, C. S., & Haggard, P. (2014). Multisensory parietal cortex contributes to visual enhancement of touch in humans: A single-pulse TMS study. *Cerebral Cortex*, *24*, 501–507.

Kube, T., Rief, W., Gollwitzer, M., Gärtner, T., & Glombiewski, J. A. (2018). Why dysfunctional expectations in depression persist—Results from two experimental studies investigating cognitive immunization. *Psychological Medicine*, *80*, 535–543.

Legrain, V., Iannetti, G. D., Plaghki, L., & Mouraux, A. (2011). The pain matrix reloaded: A salience detection system for the body. *Progress in Neurobiology*, *93*, 111–124.

Levinson, C. A., Rapp, J., & Riley, E. N. (2014). Addressing the fear of fat: Extending imaginal exposure therapy for anxiety disorders to anorexia nervosa. *Eating and Weight Disorders*, *19*, 521–552.

Lovibond, P. F., Mitchell, C. J., Minard, E., Brady, A., & Menzies, R. G. (2009). Safety behaviours preserve threat beliefs: Protection from extinction of human fear conditioning by an avoidance response. *Behaviour Research and Therapy*, *47*, 716–720.

Mayer, B., Muris, P., Kramer-Freher, N., Stout, J., & Polak, M. (2012). Covariation bias for food-related control is associated with eating disorders symptoms in normal adolescents. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, 1008–1013.

Merckelbach, H., & de Jong, P. J. (1988). Conditioned inhibition of psychophysiological responses to painful stimuli: Facts and suggestions. *Psychologica Belgica*, *28*, 139–155.

Morey, R. A., Dunsmoor, J. E., Haswell, C. C., Brown, V. M., Vora, A., Weiner, J., et al. (2015). Fear learning circuitry is biased toward generalization of fear associations in posttraumatic stress disorder. *Translational Psychiatry*, *5*, e700.

Olatunji, B. O., Cisler, J. M., Meunier, S., Connolly, K., & Lohr, J. M. (2008). Expectancy bias for fear and disgust and behavioral avoidance in spider fearful individuals. *Cognitive Therapy and Research*, *32*, 460–469.

Pauli, P., Montoya, P., & Martz, G. E. (1996). Covariation bias in panic-prone individuals. *Journal of Abnormal Psychology*, *105*, 658–662.

Peerdeman, K. J., van Laarhoven, A. I. M., Peters, M. L., & Evers, A. W. M. (2016). An integrative review of the influence of expectancies on pain. *Frontiers of Psychology*, *7*, 1270.

Proulx, T., Sleegers, W., & Tritt, S. M. (2017). The expectancy bias: Expectancy-violating faces evoke earlier pupillary dilation than neutral or negative faces. *Journal of Experimental Social Psychology*, *70*, 69–79.

Rachman, S. (1990). *Fear and courage* (2nd ed). New York: Freeman.

Remue, J., Vanderhasselt, M. A., Baeken, C., Rossi, V., Tullo, J., & De Raedt, R. (2016). The effect of a single HF-rTMS session over the left DLPFC on the physiological stress response as measured by heart rate variability. *Neuropsychology*, *30*, 756–766.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, *27*, 611–647.

Rief, W., & Glombiewski, J. A. (2016). Expectation-focused psychological interventions (EFPI). *Verhaltenstherapie*, *26*, 47–54.

Rief, W., Glombiewski, J. A., Gollwitzer, M., Schubo, A., Schwarting, R., & Thorwart, A. (2015). Expectancies as core features of mental disorders. *Current Opinion in Psychiatry*, *28*, 378–385.

Sebastiani, L., D'Alessandro, L., & Gemignani, A. (2014). Does fear expectancy prime fear? An autonomic study in spider phobics. *International Journal of Psychophysiology*, *91*, 178–185.

Seligman, M. E. P., Maier, S. F., & Solomon, R. L. (1971). Unpredictable and uncontrollable aversive events). In F. R. Brush (Ed.), *Aversive conditioning and learning* (pp. 347–400). London: Academic Press.

Sloan, T., & Telch, M. J. (2002). The effects of safety-seeking behavior and guided threat reappraisal on fear reduction during exposure: An experimental investigation. *Behaviour Research and Therapy*, *40*, 235–251.

Sussman, T. J., Jin, J., & Mohanty, A. (2020). The impact of top-down factors on threat perception biases in health and anxiety. In T. Aue & H. Okon-Singer (Eds.), *Cognitive biases in health and psychiatric disorders: Neurophysiological foundations* (pp. 215–241). San Diego: Elsevier.

Tomarken, A. J., Sutton, S. K., & Mineka, S. (1995). Fear-relevant illusory correlations: What types of associations promote judgmental bias? *Journal of Abnormal Psychology*, *104*, 312–326.

Tuominen, L., Boeke, E., DeCross, S., Wolthusen, R. P., Nasr, S., Milad, M., et al. (2019). The relationship of perceptual discrimination to neural mechanisms of fear generalization. *Neuroimage*, *188*, 445–455.

van Overveld, M., de Jong, P. J., Huijding, J., & Peters, M. L. (2010). Contamination and harm relevant UCS-expectancy bias in spider phobic individuals: Influence of treatment. *Clinical Psychology and Psychotherapy*, *17*, 510–518.

van Overveld, M., de Jong, P. J., & Peters, M. (2010). Disgust and harm related UCS expectancy bias in blood fearful individuals. *Clinical Psychology and Psychotherapy*, *17*, 100–109.

Vanderhasselt, M. A., De Raedt, R., Baeken, C., Leyman, L., Clerinx, P., & D'Haenen, H. (2007). The influence of rTMS over the right dorsolateral prefrontal cortex on top-down attentional processes. *Brain Research*, *1137*, 111–116.

vanOyen Witvliet, C., & Vrana, S. R. (2000). Emotional imagery, the visual startle, and covariation bias: An affective matching account. *Biological Psychology*, *52*, 187–204.

Verwoerd, J., de Jong, P. J., Wessel, I., & van Hout, W.J.P.J. (2013). "If I feel disgusted, I must be getting ill": Emotional reasoning in the context of contamination fear. *Behaviour Research and Therapy*, *51*, 122–127.

Vlaeyen, J. W., Crombez, G., & Linton, S. J. (2009). The fear-avoidance model of pain. *Pain*, *157*, 1588–1589.

Vroling, M. E., Glashouwer, K. A., Lange, W.-G., Allart, E., & de Jong, P. J. (2016). "What I believe is true": Belief-confirming reasoning bias in social anxiety disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *53*, 9–16.

Wiemer, J., Mühlberger, A., & Pauli, P. (2014). Illusory correlations between neutral and aversive stimuli can be induced by outcome aversiveness. *Cognition and Emotion*, *28*, 193–207.

Wiemer, J., & Pauli, P. (2016b). Enhanced functional connectivity between sensorimotor and visual cortex predicts covariation bias in spider phobia. *Biological Psychology*, *121*, 128–137.

Wiemer, J., Schulz, S. M., Reicherts, P., Glotzbach-Schoon, E., Andreatta, M., & Pauli, P. (2015). Brain activity associated with illusory correlations in animal phobia. *Social Cognitive and Affective Neuroscience*, *10*, 969–977.

Wimmer, R. D., Schmitt, L. I., Davidson, T. J., Nakajima, M., Deisseroth, K., & Halassa, M. M. (2015). Thalamic control of sensory selection in divided attention. *Nature*, *526*, 705–709.

Wood, K. H., Wheelock, M. D., Shumen, J. R., Bowen, K. H., Ver Hoef, L. W., & Knight, D. C. (2015). Controllability modulates the neural response to predictable but not unpredictable threat in humans. *Neuroimage*, *119*, 371–381.

Further reading

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, *14*, 488–501.

Wiemer, J., & Pauli, P. (2016a). How fear-relevant illusory correlations might develop and persist in anxiety disorders: A model of contributing factors. *Journal of Anxiety Disorders*, *44*, 55–62.

Wood, K. H., Ver Hoef, L. W., & Knight, D. C. (2012). Neural mechanisms underlying the conditioned diminution of the unconditioned fear response. *Neuroimage*, *60*, 787–799.