KU LEUVEN

FACULTEIT PSYCHOLOGIE EN
PEDAGOGISCHE WETENSCHAPPEN

CENTRUM LEERPSYCHOLOGIE EN EXPERIMENTELE PSYCHOPATHOLOGIE

Optimizing exposure therapy

Strengths and limitations of the extinction model

Sara Scheveneels

Proefschrift aangeboden tot het verkrijgen van de
graad van Doctor in de Psychologie

Promotor: Prof. Dr. Dirk Hermans
Copromotor: Dr. Yannick Boddez

2019

Book cover: “Spotlight” by Goran Djurovic. Special thanks to Kim Haesen.

Optimizing exposure therapy: Strengths and limitations of the extinction model

Proefschrift aangeboden tot het verkrijgen van de graad Doctor in de Psychologie door **Sara Scheveneels**

Promotor: Prof. Dr. Dirk Hermans

Copromotor: Dr. Yannick Boddez

Exposure involves the (repeated) confrontation with fear-eliciting stimuli or situations in the absence of the feared outcome and is a key-component in the treatment of clinical fears and anxiety. Several meta-analyses have demonstrated large effect sizes and response rates of exposure treatment. However, these effects are not always maintained in the long-term, leaving room for the further enhancement of clinical treatment. The first part of this dissertation (Chapter 1) introduces and evaluates fear conditioning as a laboratory model for the acquisition and treatment of clinical fears and anxiety. In addition, we discuss how recent developments in learning theory can address some of the frequently heard critiques on simple fear conditioning.

In the second part of this dissertation, we report on empirical studies that focus on optimizing exposure treatment. In Chapter 2 and 3, respectively the type and order of stimuli presented during extinction were manipulated. In Chapter 2, we investigated whether manipulating the typicality of the extinction stimulus can attenuate return of fear. We found that using a generalization stimulus during extinction that is a typical exemplar of the feared category resulted in less fear responding to a new exemplar of that feared category compared to using an atypical exemplar. In Chapter 3, we developed a fear extinction procedure to compare the outcomes of a hierarchical versus random approach of exposure. A set of morphs between the danger (CS+) and safety (CS-) cue were presented in a hierarchical order (i.e., starting with the CS-, then the morph most similar to the CS-, followed by the morph most similar to that one etc., until reaching the CS+) versus in a random order. No differences between the hierarchical and random approach were found in fear responding as tested one day later.

In Chapter 4 and 5 we explore the effects of manipulations of expectancy violation (i.e., the mismatch between the expected and actual outcome), which according to Inhibitory Learning Theory (ILT) drives inhibitory learning during exposure and long-term fear reduction. In Chapter 4, we investigated the role of expectancy violation in virtual reality exposure therapy (VRET). Both the experimental and correlational analyses failed to confirm that expectancy violation predicted treatment outcome. In Chapter 5, the effect of providing safety information before extinction – similar to types of psychoeduction before exposure – was investigated. ILT predicts that providing this type of information interferes with subsequent expectancy violation during extinction. We could not confirm that providing safety information was detrimental for the generalization of fear reduction to another context.

In the third part of this dissertation (Chapter 6), we reflect on the extinction procedures that we used in the empirical chapters and evaluate the validity of extinction research in providing insights on (how to conduct) clinical exposure therapy. Finally, our findings and conclusions are embedded in a more general discussion of some of the strengths and challenges of ILT.

Het optimaliseren van exposure behandeling: Sterktes en beperkingen van het extinctiemodel
Proefschrift aangeboden tot het verkrijgen van de graad Doctor in de Psychologie door **Sara Scheveneels**
Promotor: Prof. Dr. Dirk Hermans
Copromotor: Dr. Yannick Boddez

Exposure betreft de herhaaldelijke confrontatie met gevreesde stimuli of situaties en is een belangrijk onderdeel in de behandeling van angst. Verschillende meta-analyses bevestigen de effectiviteit van exposure. Deze effecten worden echter niet altijd op lange termijn behouden, waardoor er ruimte is om exposure behandeling in de klinische praktijk verder te optimaliseren. Het eerste deel van dit doctoraat (Hoofdstuk 1) bevat een inleiding en evaluatie van angstconditionering als een labomodel voor het verwerven en de behandeling van klinische vrees en angst. Tevens wordt beschreven hoe recente ontwikkelingen in de leertheorie een antwoord kunnen bieden op veelgehoorde kritieken op simpele angstconditioneringsmodellen.

Het tweede deel van dit doctoraat beschrijft empirische studies die focussen op het optimaliseren van exposure. In Hoofdstuk 2 en 3 wordt ingegaan op het type stimulus tijdens extinctie en de aanbiedingsvolgorde. In Hoofdstuk 2 onderzochten we of het manipuleren van hoe typerend de extinctiestimulus is voor de gevreesde categorie terugkeer van angst kan verminderen. We vonden betere generalisatie van extinctieleren naar een nieuwe stimulus uit de gevreesde categorie wanneer een extinctiestimulus werd gebruikt die typerend was voor de gevreesde categorie vergeleken met een niet-typerende extinctiestimulus. In Hoofdstuk 3 ontwikkelden we een procedure om de uitkomsten van hiërarchische versus random extinctie te vergelijken. Een set van 'morphs' tussen de bedreigende (CS+) en veilige (CS-) stimulus werd in een hiërarchische (startend bij de CS-, gevolgd door de 'morph' meest gelijkend op de CS-, enzovoort, tot de CS+) versus random volgorde aangeboden. We vonden geen verschillen tussen de hiërarchische en random benadering in een testfase die één dag later plaatsvond.

In Hoofdstuk 4 en 5 wordt ingegaan op de effecten van het manipuleren van verwachtingsdisconfirmatie (het verschil tussen de verwachte en eigenlijke uitkomst), het mechanisme dat resulteert in inhibitorisch leren tijdens exposure en angstreductie op lange termijn volgens het inhibitorisch leermodel (ILT). In Hoofdstuk 4 onderzochten we de rol van verwachtingsdisconfirmatie in virtual reality exposure therapie. Via zowel de experimentele als correlationele analyses konden we niet bevestigen dat verwachtingsdisconfirmatie voorspellend was voor de behandeluitkomst. In Hoofdstuk 5 werd het effect van psycho-educatie (met name het geven van veiligheidsinformatie) voorafgaand aan extinctie onderzocht. Op basis van ILT werd voorspeld dat het geven van dergelijke informatie de mogelijkheid om maximaal verwachtingen te disconfirmeren verstoort. We vonden echter geen nefast effect van het geven van veiligheidsinformatie voorafgaand aan extinctie op de generalisatie naar een nieuwe context.

In het derde deel van dit doctoraat (Hoofdstuk 6), reflecteren we over de extinctieprocedures die werden gebruikt in de empirische hoofdstukken en evalueren we de externe validiteit van extinctieonderzoek. Tenslotte bespreken we onze bevindingen en conclusies in het bredere kader van een aantal sterktes en uitdagingen voor ILT.

Dankuwel

aan iedereen die zich aangesproken voelt.

Table of contents

List of figures

Preface

This dissertation contains four main parts with a total of seven chapters discussing the optimization of exposure treatment for anxiety. The first part (Chapter 1) introduces fear conditioning as a laboratory model for the acquisition and treatment of clinical fears and anxiety. This theoretical chapter can be considered as background for the other (empirical) chapters of this dissertation. In the second part (Chapters 2, 3, 4, and 5) we discuss empirical work focused on optimizing the treatment of anxiety using extinction procedures as well as a clinical analogue sample. We continue with a theoretical reflection on the extinction model that we used in our empirical work and evaluate Pavlovian fear extinction as a model for clinical exposure treatment in the third part of this dissertation (Chapter 6). Each chapter contains a manuscript with a separate introduction and discussion. Therefore, this preface only aims to provide a general background together with an outline of the different chapters and how they are related.

Part 1: Fear conditioning as a model for the acquisition and treatment of clinical fear

Learning relations between stimuli is crucial in the etiology and treatment of anxiety (e.g., Mineka & Zinbarg, 2006). For example, an individual that has experienced severe turbulence during a flight, might start to react fearfully to flying. To investigate the role of learning mechanisms in anxiety in a controlled and standardized way, fear conditioning and extinction procedures have been used (e.g., Beckers, Krypotos, Boddez, Effting, & Kindt, 2013; Hermans, Craske, Mineka, & Lovibond, 2006; Lonsdorf et al., 2017). In fear conditioning, one stimulus (i.e., conditioned stimulus; CS) is (repeatedly) paired with another stimulus (i.e., unconditioned stimulus; US), resulting in changes in responding to the CS. During extinction, the CS is subsequently presented in the absence of the US, typically resulting in diminished responding to the CS. In Chapter 1 we introduce the fear conditioning model and discuss its relevance to clinical fears and anxiety. Since one of the main goals of fear conditioning research is to enhance the understanding of clinical anxiety, the question arises whether it is a valid laboratory model that is informative for clinical practice. In Chapter 1 we aim to provide an answer to this question. Guided by some often heard critiques on simple fear conditioning, we discuss how more recent developments in learning theory can address earlier shortcomings.

Part 2: Empirical work on optimizing exposure treatment for anxiety

Similar to an extinction procedure, exposure treatment involves the (repeated) confrontation with fear-eliciting stimuli or situations in the absence of the feared outcome. Exposure is a key-component in the treatment of clinical fears and anxiety, with several meta-analyses demonstrating large effect sizes and response rates (e.g., Hofmann & Smits, 2008; Loerinc et al., 2015). However, these effects are not always maintained in the long-term, leaving room for the further enhancement of clinical treatment (Craske & Mystkowski, 2006).

In the second part of this dissertation we discuss empirical studies focused on the optimization of exposure treatment. The first two studies focus on how the kind of stimuli used during extinction and the order in which the stimuli are presented influence fear reduction. In clinical practice, the original CS is often no longer available or cannot be identified and generalization stimuli (GSs) have to be used in exposure treatment. Previous research, however, shows that successful fear reduction after using a GS during extinction can be followed by a return in fear responding when being presented with the original CS or new GSs (e.g., Vervliet, Vansteenwegen, Baeyens, Hermans, & Eelen, 2005; Zbozinek & Craske, 2018). The first study (Chapter 2) investigates whether using a GS that is a typical exemplar of the feared category during extinction can attenuate return of fear to a new exemplar as compared to using an atypical extinction GS.

In addition to the kind of stimuli used during exposure, patients can be exposed to stimuli in a hierarchical way (i.e., gradually proceeding from less to more fear-eliciting stimuli) or in a random way (i.e., being exposed to less and more fear-eliciting items randomly). In the second study (Chapter 3), we developed a fear extinction procedure to compare the long-term outcomes of a hierarchical versus random approach of exposure. Based on Inhibitory Learning Theory (ILT; Craske et al., 2008; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014), it was predicted that random extinction would result in lower long-term fear responding than hierarchical extinction. In particular, ILT argues that this effect might be driven by random extinction allowing for higher expectancy violation or a mismatch between the expected feared outcome and the actual outcome. According to ILT the larger this mismatch, the more inhibitory learning can occur (Rescorla & Wagner, 1972). Such inhibitory learning (i.e., learning a CS-noUS association) is assumed to counteract on the existing (excitatory) CS-US association, resulting in extinguished fear responding.

We further explore the role of expectancy violation in exposure treatment in Studies 3 and 4, respectively Chapters 4 and 5 of this dissertation. As mentioned earlier, according to the guidelines rooted in ILT, exposure treatment will result in better long-term fear reduction if it allows for maximal expectancy violation. This implies that an important task of the therapist is to identify exactly what outcome(s) the patient fears the most (i.e., the US) and provide exposure to those situations in which the expectancy of this most feared outcome(s) is maximal. It can be argued that virtual reality exposure therapy (VRET) is not optimal for this purpose, since (at least some) feared outcomes cannot occur in a virtual environment. For instance, severe injury or death by a plane crash (in fear of flying) or a biting incident (in fear of dogs or of other animals) is unlikely to occur in VRET. Paradoxically, VRET has found to be as effective as exposure in vivo in some studies (e.g., Opriş et al., 2012; Powers & Emmelkamp, 2008). Study 3 (Chapter 4) investigates the role of expectancy violation in VRET in an experimental and correlational way. In a clinical-analogue sample consisting of students anxious of public speaking, it was experimentally manipulated whether expectancies with regard to the (overt) reactions of the audience could be violated during VRET by instructing participants whether or not the reactions of the audience where adjusted to their presentations. In addition, we asked each participant individually to report his expectancies in public speaking situations and indicate which of these expectancies were testable in the VRET. We examined whether individual differences in the proportion of expectancies that were indicated as being testable in VRET predicted treatment outcome.

Although the exposure treatment itself might allow for expectancy violation, it is possible that certain interventions of the therapist before the exposure treatment attenuate the patient's feared expectancies. This might, in particular, be the case for types of psychoeducation in which the patient is provided with information about the (low) probability that the feared outcome would occur. In flying phobia, for instance, the (extremely low) probability of a plane crash can be discussed with patients to lower the threshold to engage in the in vivo exposure (i.e., taking a plane). However, it has been argued that providing such safety information beforehand interferes with the possibility to maximally violate expectancies in subsequent exposure treatment and would therefore be deleterious for the effects of the exposure (Craske et al., 2014). If a patient is already told that the probability of his feared outcome is (extremely) low, he or she might indeed be not as surprised as without having received this safety information if the feared outcome does not occur during exposure treatment. In Study 4 (Chapter 5) we

investigate the effects of providing such safety information on return of fear in a fear conditioning procedure.

Part 3: The validity of extinction as a laboratory model for exposure treatment

In three out of four empirical studies that are discussed in the second part of this dissertation, we used extinction paradigms in the laboratory to provide insights in how to conduct clinical exposure treatment. We used these procedures based on the common assumption that Pavlovian fear extinction is a valid model for exposure. In the third part of this dissertation (Chapter 6), we critically and systematically evaluate this assumption, using three established criteria of external validity: face validity, construct validity and predictive validity (Davey, 2017; Luyten, Vansteenwegen, van Kuyck, Gabriëls, & Nuttin, 2011; Vervliet & Raes, 2013). In addition, we demonstrate how adjustments to the fear extinction model might increase its external validity.

Finally, we end this dissertation with a summary and general discussion of our findings, embedded in a broader reflection on the strengths of and challenges for ILT.

References

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

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

Craske, M. G., & Mystkowski, J. (2006). Exposure therapy and extinction: Clinical studies. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and learning: Basic science to clinical application* (pp. 213-233). Washington, DC: American Psychological Association.

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. doi:10.1016/j.brat.2014.04.006

Davey, G. C. (2017). A research pathway for experimental psychopathology: the role of external validity criteria. *Psychopathology Review, 4*, 129-140.

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

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *The Journal of Clinical Psychology, 69*, 621-632.

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

Lonsdorf, T., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., …, & Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience and Biobehavioral Reviews, 77*, 247-285. doi:10.1016/j.neurbiorev.2017.02.026

Luyten, L., Vansteenwegen, D., van Kuyck, K., & Nuttin, B. (2011). Contextual conditioning in rats as an animal model for generalized anxiety disorder. *Cognitive, Affective, & Behavioral Neuroscience, 11*, 228-244. doi:10.3758/s13415-011-0021-6

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology

of anxiety disorders. *American Psychologist, 61*, 10-26. doi:10.1037/0003-066X.61.1.10

Opriş, D., Pintea, S., García-Palacios, A., Botella, C., Szamosközi,, Ş., & David, D. (2012). Virtual reality exposure therapy in anxiety disorders: A quantitative meta-analysis. *Depression and Anxiety, 29*, 85-93. doi:10.1002.da.20910

Powers, M. B., & Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorder: A meta-analysis. *Journal of Anxiety Disorders, 22*, 561-569. doi:10.1016/j.anxdis.2007.04.006

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

Vervliet, B., & Raes, F. (2013). Criteria of validity in experimental psychopathology: Application to models of anxiety and depression. Psychological Medicine, 43, 2241 -2244. doi:10.1017/S0033291712002267

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy, 43*, 357-371. doi:10.1016/j.brat.2004.02.005

Zbozinek, T. D., & Craske, M. G. (2018). Pavlovian extinction of fear with the original conditional stimulus, a generalization stimulus, or multiple generalization stimuli. *Behaviour Research and Therapy, 107*, 64-75. doi:10.1016/j.brat.2018.05.009

Chapter 1

Learning mechanisms in fear and anxiety: It is still not what you thought it was

Based on:

Scheveneels, S., Boddez, Y., & Hermans, D. (2018). Learning mechanisms in fear and anxiety: It is still not what you thought it was. In B. Olatunji (Ed.), *The Cambridge Handbook of Anxiety and Related Disorders.*

Abstract

This chapter describes the role of learning processes in fear and anxiety. After introducing the fear conditioning model, we evaluate the external validity of this model. This relates to the question of whether fear conditioning experiments can inform clinical practice. Although there are clear similarities between the fear conditioning model and clinical anxiety at face value, some important critiques have been formulated with regard to whether the same etiological processes are at play. Specifically, early learning models provided an overly simplistic view on anxiety disorders and are insufficient to explain why some individuals develop an anxiety disorder and others do not. Extending the seminal work of Mineka and Zinbarg (2006), we demonstrate how contemporary learning theory can overcome these critiques. In addition, complex learning procedures are described that better mimic the etiological processes at play in anxiety disorders. In the last section, we describe evidence with regard to the predictive value of the conditioning model. In particular, we discuss environmental and individual-level factors that exert a similar influence in the conditioning model and in real life, demonstrating that the fear conditioning model allows for translation to real-life situations that pertain to clinical anxiety.

Keywords: Learning, Conditioning, Fear, Anxiety, External Validity

Learning mechanisms in fear and anxiety: It is still not what you thought it was

Alex is a 24-year old male who recently dropped out of work because it became difficult to function well in his job as a communication-assistant. For example, he avoided making telephone calls in the presence of others and avoided having lunch with his colleagues. When talking to a superior in particular, Alex is extremely self-conscious and aware of certain physical symptoms such as blushing, sweating and trembling. He is afraid that others will notice these symptoms and would evaluate him as being incompetent, stupid or as being a weirdo. Alex grew up in a family of lawyers, in which the importance of a professional career was stressed very much. Alex mentions that he has always been timid, but his social anxiety became worse at university, and in particular after he gave a wrong answer on a question during class. The professor reacted by saying that he does not belong at university if he cannot answer a question which is that obvious. Alex reacted to this by running out of the classroom. He remembers that his classmates were laughing when he was running out of the room, but he is not sure whether this memory is accurate or whether he imagined it. After this incident, Alex started to skip classes and avoided contact with his classmates.

Learning experiences play a crucial role in the etiology of anxiety. An individual with social anxiety, such as in the example of Alex, might have learned that (saying something stupid in) the company of others is related to rejection and exclusion, which might have caused the current symptoms of social anxiety and avoidance to manifest itself. Learning theory has not only provided valuable insights into the onset of anxiety disorders, it has also provided and still does provide a major impetus in the development and optimization of the treatment of anxiety.

In this chapter we will discuss how learning mechanisms as investigated in basic fear conditioning research are at work in clinical fear and anxiety. We will mainly discuss research in human (healthy and anxious) participants and occasionally discuss findings from rodent research as well. The first section of this chapter will focus on different procedures and outcomes when modeling fear acquisition in the laboratory. Subsequently, we will evaluate whether we can translate the findings from fear conditioning research in the laboratory to clinical anxiety based on three established criteria of external validity: face validity, construct validity and predictive validity. We will demonstrate that recent developments in the field can respond to often heard critiques on the classical fear conditioning model of anxiety.

A learning perspective on the etiology of anxiety disorders

A classical conditioning account of anxiety

One of the first and without doubt most famous case studies illustrating that fear reactions can be acquired by learning experiences is the one of Albert B. as described by Watson and Rayner (1920). A phobic reaction was induced in Albert by pairing a stimulus that initially did not evoke fear (i.e., a white rat) with an aversive outcome (i.e., a loud noise). After repeatedly presenting the rat together with the loud noise, Albert started to react fearfully to the rat. This is one of the first demonstrations of fear conditioning, nowadays an established procedure to induce fear in the laboratory.

The early demonstration of Watson and Rayner that anxiety for a stimulus can develop by learning experiences, and in particular by pairing the stimulus with an aversive event, relies on the experimental work of Pavlov (1927). Using dogs as subjects, Pavlov paired stimuli such as sounds with food intake. Pavlov demonstrated that the dogs initially did not show increased saliva production in response to the sound. However, after several pairings of the sound together with the food, the sound started to elicit increased saliva production. This is referred to as *classical or Pavlovian conditioning*. We will now use Pavlov's procedure to define the concepts involved in learning. The increased saliva production in response to the sound (CS) is called the *conditional response* (CR). A CR can be defined as a change in responding that is conditional upon the relation between the presence of at least two stimuli (De Houwer, Barnes-Holmes, & Moors, 2013). In Pavlov's procedure, one of these two stimuli is the sound, which functions as a *conditional stimulus* (CS). Such CS can be defined as a stimulus to which responding changes conditional upon a relation with another stimulus (i.e., with the food). The food stimulus itself is termed the *unconditional stimulus* (US). In case of successful conditioning, the US changes responding to the stimulus it is related to (i.e., to the CS)[1]. Notably, in Pavlov's early experiments the US was an appetitive stimulus (i.e., food). This is different for fear conditioning research, where the US is an aversive stimulus (e.g., loud noise). However, as will become clear in the next paragraph, besides this, the procedure is fairly similar.

[1] Changes in responding to the US, as caused by its relation with the CS, are possible but typically left uninvestigated in conditioning research.

(Human) fear conditioning: Procedures and outcomes

The fear conditioning model is widely recognized as a model for the pathogenesis of fear and anxiety disorders (e.g., Beckers, Krypotos, Boddez, Effting, & Kindt, 2013). Typically, the CS is paired (repeatedly) with an aversive US until it elicits conditional responding indicative of fear and anxiety.

In human fear conditioning a variety of stimuli has been used as a *CS*, ranging from fairly neutral stimuli (e.g., geometrical shapes) to fear-relevant stimuli (e.g., spiders, fearful faces). Often the CS is a visual stimulus, but other modalities such as auditory and olfactory stimuli have been used as well. An electrocutaneous stimulus or electric shock is typically used in the laboratory as a *US*, with an intensity set at a level which is perceived as 'uncomfortable, but not painful' by the individual participant. However, also auditory stimuli such as human screams or (bursts of a) loud (white) noise (100-105 dB) and aversive pictures or movie clips have been used as a US (e.g., Joos, Vansteenwegen, & Hermans, 2012; Lenaert et al., 2014). For modelling panic symptoms in the laboratory, CO_2- enriched air leading to a sensation of breathing restriction has proven successful (Leibold et al., 2013).

As described earlier, after pairing a CS (repeatedly) with a US, presenting the CS will elicit a *CR* indicative of fear or anxiety. Several dependent variables can be included as indices of fear and anxiety in response to the CS. In line with Lang's bio-informational theory (1979) and emotion theory (Frijda, 1986) dependent variables in fear conditioning research can focus on each of the three response systems: the verbal, psychophysiological, and behavioral level.

On the *verbal* level, US-expectancy ratings are commonly used by asking participants to indicate the extent to which they expect US-occurrence (Boddez et al., 2013). In the example of Alex, this could correspond to the extent to which he expects that others will exclude or reject him after saying something (stupid) during lunchbreak. In addition to US-expectancy ratings, ratings of subjective fear, CS valence and subjective units of distress are sometimes included.

As *physiological* indices, skin-conductance response (SCR) and fear-potentiated startle (FPS) have a long history in human fear conditioning research. In SCR (also known as galvanic skin response, electrodermal responding or sympathetic skin response) changes in the electrical conductance of the skin are measured. This is informative for fear learning because the conductive properties of the skin are influenced by sympathetic autonomic arousal and are reactive to signals of salient events. Notably, SCR is not specific for the anticipation of aversive events, but increases in response to any salient event (e.g., also appetitive ones; Lipp, 2006). In

contrast to SCR, a potential advantage of FPS is that it is modulated by emotional valence (Grillon & Baas, 2003, but see Mallan, Sax, & Lipp, 2009). For measuring FPS in humans, an eye-blink reflex is elicited by the administration of a high-intensity probe of white noise (e.g., Blumenthal et al., 2005). The amplitude of this eye-blink reflex, as measured by the electromyographic activity of the muscles around the eye (orbicularis oculi), is potentiated in anticipation of aversive events. In addition to SCR and FPS, other physiological indices of fear and anxiety, such as heart rate and pupil dilation have been included in human fear conditioning studies (e.g., Leuchs, Schneider, Czisch, & Spoormaker, 2017). Importantly, these physiological indices correspond to the physical symptoms in clinical anxiety. In the clinical example, Alex started to sweat when he was confronted with a fear-eliciting situation such as when talking to a superior, typically resulting in increased electrical conduction of the skin. In addition, his heart started pounding really fast. He also reported that, while being at the office, he had a tendency to startle each time the telephone of a colleague ringed.

A third set of dependent variables is situated on the overt behavioral level. As a key characteristic and one of the diagnostic criteria of anxiety disorders (American Psychiatric Association, 2013), *avoidance behavior* can be considered an indispensable outcome measure when modeling pathological anxiety. Alex' avoidance behavior was debilitating and prevented him to do his job appropriately. Because he was afraid of saying something stupid or making a bad impression, he avoided making telephone calls in the presence of colleagues, going to team meetings and having lunch with colleagues or superiors. Nevertheless, whereas in most fear conditioning studies typically at least one physiological and one verbal measure is included for reasons of convergent validity, only a few research groups consistently include the assessment of avoidance (tendencies) (Beckers et al., 2013). In fear conditioning research, avoidance can be measured by giving participants the option to avoid the CS (e.g., Grillon et al., 2006) or to avoid the US when presented with the CS (e.g., van Meurs, Wiggert, Wicker, & Lissek, 2014). Moreover, the avoidance response can vary in cost, ranging from button presses to avoid the US without additional response cost (Lovibond, Mitchell, Minard, Brady, & Menzies, 2009; Vervliet & Indekeu, 2015) to avoidance responses associated with a cost, for instance a loss in points (Pittig, Brand, Pawlikowski, & Alpers, 2014). In addition, visual avoidance (e.g., looking away) can be assessed by using eye-tracking methodology (e.g., Rinck & Becker, 2006). Some studies include a separate approach-avoidance task to assess avoidance tendencies or the urge to avoid (e.g., Krypotos, Effting, Arnaudova, Kindt, & Beckers, 2014; Krypotos, Arnaudova, Effting, Kindt, & Beckers, 2015, Van Gucht, Vansteenwegen, Van den Bergh, & Beckers,

2008). In such task, a manikin figure typically appears at the bottom or top of the screen and the CS is presented in the opposite half of the screen. Participants are instructed to move the manikin towards (approach) or away from (avoidance) the CS as quickly as possible by pressing a key or by handling a joystick. The time between the CS onset and the response is measured. If participants are faster to move the manikin away from the CS than towards it, a tendency to avoid the CS is inferred.

For reasons of cross-validation, fear conditioning studies typically include more than one dependent variable, usually a verbal measure and one or more psychophysiological measures. Importantly, correlations between outcome measures often have shown to be only weak (e.g., Hodgson & Rachman, 1974). This observed divergence in outcome measures can, apart from measurement error, also be explained by the fact that the three response systems indeed represent different and partly independent dimensions of fear and anxiety (Beckers et al., 2013).

An important question is whether (human) fear conditioning can indeed serve as a model for clinical anxiety disorders. As we discussed in the previous paragraph, at a procedural level and with regard to outcome measures the fear conditioning model shows some clear similarities with the pathogenesis and symptomatology of clinical anxiety. In the next paragraph, we will evaluate the validity of human fear conditioning as a model for clinical anxiety disorders in a more systematic way.

The external validity of the (human) fear conditioning model

It is considered one of the most important advantages of (human) fear conditioning research that it allows modeling pathological behavior in healthy subjects. A main assumption is that the knowledge and insights acquired through highly controlled fear conditioning experiments can inform clinical practice. This assumption concerns the external validity of the fear conditioning model. In the past, several critiques have been formulated on this model, arguing that it might be too simplistic to capture the complexity of clinical anxiety. Mineka and Zinbarg (2006) replied to these critiques and illustrated how contemporary approaches of conditioning and learning theory can overcome a substantial part of them. Below, we further elaborate on the work of Mineka and Zinbarg (2006) by providing a systematic evaluation of the fear conditioning model using three established criteria of validity: face validity, construct validity

and predictive validity[2]. These criteria have been used to evaluate external validity in pharmacological and more recently in behavioral research (Davey, 2017; Luyten, Vansteenwegen, van Kuyck, Gabriëls, & Nuttin, 2011; Scheveneels, Boddez, Vervliet, & Hermans, 2016; Vervliet & Raes, 2013).

Face validity. The most straightforward validity criterion is face validity. This refers to the surface similarity between the (fear conditioning) model and the (anxiety) disorder. As discussed in the previous section, at face value, the fear conditioning model shows important similarities with clinical anxiety. Conditional reactions that are observed in fear conditioning studies such as SCR, startling, elevated heart rate, correspond to the physical reactions in real-life fear-eliciting situations. In addition, (a tendency) to avoid the conditional stimulus as well as the expectancy that the aversive outcome will occur can be observed in both fear conditioning studies and clinical anxiety.

Notably, fear conditioning research in the laboratory typically uses simple, unimodal stimuli such as geometrical shapes, whereas in real-life the learning experiences often take place in more complex circumstances. Therefore, more recently, the use of complex, multi-sensory stimulus configurations (e.g., auditory, tactile, olfactory, visual) has been proposed to provide a better analogue for real-life experiences (Waters, LeBeau, & Craske, 2017). Virtual and augmented reality might be a useful technology to achieve this (e.g., Baas, Nugent, Lissek, Pine, & Grillon, 2004). However, the added value of using these more complex stimulus configurations still requires empirical testing. More specifically, future research should reveal whether using more complex stimulus configurations (as compared to simple stimuli) allows for better translation to clinical anxiety with respect to, for example, newly proposed treatment strategies.

Establishing and optimizing the face validity of the fear conditioning model can be considered a first step, but it is not the most decisive criterion of external validity. In the next paragraph we will continue with evaluating more significant criteria.

Construct validity. The evaluation of the construct validity revolves around the question whether the same (etiological) processes are at play in the model and the clinical disorder.

[2] The reader may note that the subtitle of our chapter ("It is still not what you thought it was") is similar to the subtitle used in the article of Mineka and Zinbarg (2006).

Different from face validity, construct validity is not about a first impression, but about underlying mechanisms.

A first indication for similar (neurobiological) processes underlying the fear conditioning model and clinical anxiety can be found in neuroimaging studies that reveal a similar neurocircuitry involved in anxiety patients and in fear conditioning in animals and healthy humans (e.g., Sehlmeyer et al., 2009). In particular, a central role of amygdaloid nuclei has consistently been demonstrated in the acquisition and expression of fear responses and across different anxiety disorders (e.g., Kent & Rauch, 2003; Sehlmeyer et al., 2009; Shin & Liberzon, 2010). In addition, other brain areas such as the hippocampus, insula, anterior cingulate cortex and ventromedial prefrontal cortex have been identified as regions of interest across anxiety disorders and in fear conditioning research (e.g., Damsa, Kosel, & Moussally, 2009; Etkin & Wager, 2007; Ipser, Singh, & Stein, 2013)[3].

An assumption of the conditioning approach is that learning processes underlie the etiology and treatment of clinical anxiety. Early learning theorists such as Watson used simple acquisition procedures to model the etiology of anxiety disorders. In these procedures one neutral stimulus (e.g., a white rat) was paired with one aversive outcome (e.g., a loud noise). This simple acquisition procedure has been subject to some important critiques. We will discuss some of these earlier shortcomings as well as how they are addressed by more recent developments in learning theory.

A contemporary learning theory approach on the etiology of anxiety disorders. One criticism that has been formulated on the conditioning perspective is *that many anxiety patients are not able to report a CS-US event that can account for the current anxiety symptoms*. A first possibility to nonetheless explain this from a learning account is to simply assume that a CS-US event has occurred, but that people are not accurate in remembering and retrospectively reporting this conditioning experience (Merckelbach, van den Hout, Hoekstra, & de Ruiter, 1989; Öst & Hugdahl, 1981).

Alternatively and maybe more likely, generalization and higher order conditioning might obscure learning as the actual cause of the anxiety symptoms. Generalization will be discussed in more detail below but can already be illustrated by an anecdote about Albert from the study

[3] Results regarding these areas are less univocal and their role remains subject of discussion (e.g., Maren, 2008; Sehlmeyer et al., 2009).

by Watson and Rayner (1920). It was said that Albert did not only fear white rats after the conditioning experience, but also women wearing fur coats. If Albert would have entered treatment with this complaint, one might have speculated about the sensational (Freudian) origins of this fear. Nonetheless, it can easily be explained by a learning account if one assumes that the fear generalized from the rat to fur coats based on their looking alike. With respect to higher order conditioning, emetophobia (i.e., fear of vomiting) may serve as an example. Although some dry foods (e.g., cereal without milk) might never have been directly involved in an aversive learning experience, the patient might relate these foods to difficulties to swallow, which in turn is related to vomiting (Bouman & van Hout, 2006). This allows to explain fear and avoidance of these foods from a learning perspective.

As discussed above, conditioning can be defined as a change in responding due to a relation in the presence of stimuli (De Houwer et al., 2013). Importantly, this definition also covers observational learning (e.g., Cameron, Roche, Schlund, & Dymond, 2016). Take the example of a patient who developed dog phobia after observing somebody else get bitten by a dog. Crucially, in such case, the fear can still be explained by (an observed) relation between the presence of stimuli; more precisely, by a relation between the presence of a dog and a dog bite. In addition, one might develop dog phobia if one is told that dogs cause bit wounds. In such fear learning via instruction, the fear is caused by a verbal description of the relation between the presence of stimuli. Accounting for learning via observation and instruction of course significantly increases the explanatory territory of the learning framework. Olsson and Phelps (2004) compared fear acquisition through direct pairings, instructions and observational learning in an experimental study with human participants. In the Pavlovian (direct) learning group, a CS was paired with an electric shock. The observational-learning group observed a confederate that was presented with the CS paired with shock. Participants in the instructed-learning group were given verbal instructions about the CS being followed by shock. Interestingly, similar levels of conditional fear responding, as measured by the skin-conductance response, were found for all three pathways. Moreover, in follow-up studies, it has been demonstrated that the neural correlates of direct and indirect pathways to fear acquisition are largely overlapping, with amygdala activation observed in both pathways (Olsson, Nearing, & Phelps, 2007; Phelps, Connor, Gatenby, Gore, & Davis, 2001). These results provide experimental evidence that learning processes can be involved even when no direct conditioning experience is reported by the patient (also see Rachman, 1977).

Another reason why patients may not be able to report a CS-US event that can account for their anxiety symptoms is that USs may be more subtle than the bite of an animal. For example, in patients suffering from chronic fatigue syndrome something as subtle as experiencing fatigue may function as a US. If such patients learn a relation between stair climbing and fatigue, stair climbing may come to elicit fear (Lenaert, Boddez, Vlaeyen, & van Heugten, 2018). In the case of such subtle USs, it is easy to overlook that learning is involved. A similar argument holds for other interoceptive USs like panic (Bouton, Mineka, & Barlow, 2001).

Until now, we discussed how learning theory can handle the observation that many anxiety patients are not able to report a CS-US event that can account for their anxiety symptoms. Another intriguing observation which learning theory has to account for is that *not everyone undergoing a traumatic event will eventually develop an anxiety disorder* (Poulton & Menzies, 2002). Although about 95% of people experience one or more traumatic events during their lifetime, only 10-30% develops an anxiety disorder (Engelhard, van den Hout, & McNally, 2008). One can make this insightful by assuming that there are additional variables that moderate the learning process. Such variables include *temperamental/biological vulnerabilities* as well as inter-individual differences in *contextual/experiential factors before, during and following the conditioning experience*. In the following paragraphs a selection of these variables will be discussed. For a more comprehensive overview, we refer the interested reader to Lonsdorf and Baas (2015) and Lonsdorf and Merz (2017).

Given the same conditioning experience, some individuals are more vulnerable to develop an anxiety disorder than others, due to moderation by individual differences in genetic predisposition and temperamental factors. So far, among the most established genetic factors identified as being related to anxiety problems, is a polymorphism in the serotonin transporter gene promotor region, 5-HTTLPR[4] (e.g., Lonsdorf & Kalisch, 2011). In particular, carriers of the 5-HTTLPR s-allele have found to suffer from more severe panic symptoms and social anxiety (Lonsdorf et al., 2009; Miu, Vulturar, Chis, Ungureanu, & Gross, 2013; but see Blaya, Salum, Lima, Leistner-Segal, & Manfro, 2007). The 5-HTTLPR s-allele has found to be associated with anxiety-related personality traits as well (Munafò et al., 2009). In addition, it has been demonstrated consistently that the low-efficacy 5-HTTLPR s-allele is associated with

[4] Importantly, the field of genetics in psychopathology is characterized by several pitfalls and limitations such as small effects, post-hoc testing, underpowered studies and failed replications. We refer the interested reader to Dick et al., (2015), Duncan and Keller (2011) and Tabor, Risch, and Myers (2002) for a more elaborate discussion of this topic.

facilitated fear conditioning (e.g., Bauer, 2015; Lonsdorf et al., 2009; Wendt et al., 2014). Importantly, in line with a diathesis-stress model of psychopathology, 5-HTTLPR genotype has found to interact with stressful life events. Klucken et al. (2012), for example, found that carriers of the 5-HTTLPR s-allele showed elevated activity in the neural fear network in response to a CS, but only if they had a history of stressful life events.

In addition, particular *personality traits* might moderate the relation between learning and anxiety. An extensive set of personality traits has been investigated in this context. Among the most common studied measures is *intolerance of uncertainty (IU)*. Individuals scoring high on IU react negatively to uncertain situations and consider such situations as threatening (e.g., Carleton, 2016; Lonsdorf & Merz, 2017). Available evidence shows that IU is a transdiagnostic risk factor for the development and maintenance of anxiety disorders (e.g., Gentes & Ruscio, 2011; McEvoy & Mahoney, 2012). In some fear conditioning studies, IU has been found to be positively correlated with fear responding and generalization (e.g., Chin, Nelson, Jackson, & Hajcak, 2016; Morriss, Macdonald, & van Reekum, 2016; Nelson, Weinberg, Pawluk, Gawlowska, & Proudfit, 2015). Notably, these results were only observed under conditions in which the aversive US (i.e., an electric shock) followed on 50% of the trials (Chin et al., 2016). These results confirm that individuals high on IU show stronger fear responding in ambiguous situations. Other studies, however, have failed to replicate these findings (e.g., Arnaudova et al., 2013).

In addition, contextual factors *before*, *during*, or *after* the crucial conditioning event can moderate the outcome of it by either protecting the individual against developing an anxiety disorder or by making the individual more vulnerable. Therefore, it seems important to take into account the entire learning history when trying to explain why some individuals do and some do not develop an anxiety disorder even though experiencing the same trauma. We will now demonstrate that these contextual factors fall within the scope of learning theory as well.

If an individual has safe experiences with the CS *prior to* conditioning, this can attenuate the development of a CR. For example, if a tone is first repeatedly presented by itself and in a second phase presented together with electric shock, then fear will develop slower as compared to when pre-exposure to the tone did not happen. In learning theory, this phenomenon is referred to as latent inhibition (Lubow & Moore, 1959). Similarly, it has been shown that non-traumatic experiences with the event or stimulus prior to trauma (e.g., visiting the dentist) is protective against the development of psychopathology (e.g., dental phobia) (Kent, 1997). On the other hand, pre-existing stress or trauma can make an individual more vulnerable for developing an

anxiety disorder following a conditioning event. In a study by Rau, Decola, and Fanselow (2005), rats were exposed to a very mild electric shock in a specific cage (say cage B). Importantly, the experimental group received 15 heavy shocks in a very different cage in advance (say cage A), whereas the control group did not undergo this additional treatment. At test, rats in the experimental group behaved very fearful in cage B (more so than rats in the control group), even though they had only received a mild shock in this cage. This might serve to explain why posttraumatic stress disorder (PTSD) patients or individuals with already increased stress levels are more sensitive to develop new anxieties after a mild aversive event. In the example of Alex, having bad experiences in social situations in the past (e.g., being bullied as a child) could have made him more vulnerable to develop social anxiety after the incident with the professor.

During conditioning or the traumatic event, having a sense of control or mastery might be a protective factor. Mineka, Cook, and Miller (1984) presented one group of rats with unsignaled escapable shocks and another group of yoked subjects with the same amount of unsignaled but – crucially – inescapable shocks. The group that could not escape the shocks showed more freezing to both the conditioning context and cue than the group that could escape the shock. Similar results have been found in humans by Meulders et al. (2012). Participants that had control over the offset of the US showed less conditional responding than a group of participants that had no such control. Almost immediately after Alex gave the wrong answer, he realized that he had misunderstood the professor's question. Alex could have felt more control or mastery if he had replied to the professor's nasty remark with a joke or even with a simple statement that he had misunderstood the question. Instead, he panicked and ran out of the room.

Finally, contextual variables *after* conditioning can moderate whether or not a learning experience results in clinical anxiety. In learning theory, a phenomenon referred to as US-inflation has been described: the isolated presentation of a strong US after a conditioning experience with a mild US can result in an increase in conditional fear responding (Rescorla, 1974). For example, having a severe panic attack after having experienced mild panic in an elevator might strengthen fear responding to the elevator even if the severe panic attack did not occur in the elevator. Importantly, the inflation effect can also be installed by providing verbal information about the US (e.g., den Hollander, Meulders, Jakobs, & Vlaeyen, 2015). For instance, an individual can initially experience only limited driving anxiety and avoidance after a car accident. However, verbal threat information about the potential consequences of the

accident by others afterwards might increase fear responding and avoidance. Similarly, Alex' parents, attaching much importance to a professional career, responded to the incident by saying that he will definitely not be able to pursue a career at a university following such negative remark from the professor.

In addition, repeatedly thinking or ruminating about the traumatic event can make an individual more vulnerable for developing an anxiety disorder. In particular, worrying or ruminating about the conditioning experience might result in repeated activation of the CS-US contingency which might strengthen and retain the fear memory. Joos, Vansteenwegen, and Hermans (2012) investigated this in a human fear conditioning study. In an acquisition phase, participants learned two CS-US contingencies. Both CSs were pictures of faces, the USs were a human scream and a burst of white noise. Participants were instructed to rehearse one of these contingencies in an experimental session and during the subsequent week. More precisely, they were asked to "think back to the picture (CS), the scream/noise (US) and the relationship between them". When tested with both CSs one week after acquisition, it was found that fear responding was better retained for the contingency that was rehearsed than for the non-rehearsed contingency. Importantly, these results could not be explained by merely rehearsing the CS or by increasing the negative value of the US representation (i.e., US-inflation; Davey & Matchett, 1994). In a follow-up study by Joos, Vansteenwegen, Vervliet, and Hermans (2013), rehearsal of the CS alone failed to produce sustained responding. Moreover, in Joos et al. (2013) two CSs were paired with the same US. Whereas changes in the US-representation would impact both CSs similarly, it was found that only the CS from the rehearsed CS-US contingency resulted in sustained responding. This indicates that the CS, US and the contingency between them have to be rehearsed in order to obtain the effect. Taken together, this set of experiments provides another explanation of why not everyone undergoing a traumatic event will eventually develop an anxiety disorder: rumination moderates the outcome of the learning process, in such way that rumination increases the risk.

More complex procedures to model the acquisition of fear and anxiety. Early learning theorists such as Watson used simple acquisition procedures, in which one neutral stimulus was paired with one aversive outcome, to model the etiology of anxiety disorders. This simple acquisition procedure has been subject to some important critiques. In the previous section, we discussed how a learning theoretical account can overcome this criticism and expand its explanatory territory by including phenomena that moderate the learning process. In this

section, we will introduce a set of complex learning procedures that further add to the construct validity of the fear conditioning approach. More precisely, we will discuss three procedures that mimic important but underappreciated processes underlying anxiety disorders: (1) context conditioning, (2) inhibitory conditioning, and (3) generalization[5].

In a typical *context conditioning* procedure, a CS is paired with a US (i.e., simple acquisition procedure) in a control group, whereas the experimental group receives USs that are presented explicitly unpaired with the CS (e.g., Grillon & Davis, 1997). As a result of these unpaired CS-US presentations, the context (rather than a discrete CS) indicates that the US might occur somewhere in the not too distant future. Note that context in these experiments is typically operationalized as a background picture or color on the computer screen that stretches out in time before and after CS and US presentation (although other operationalizations have been used as well; Boddez et al., 2014). Context conditioning typically results in a sense of unpredictability and in a generalized and sustained state of arousal, because participants cannot precisely predict when the US will occur (e.g., Grillon, 2002; Grillon, Baas, Lissek, Smith, & Milstein, 2004).

Context conditioning has been proposed as a model for pathological conditions that are characterized by chronic, future-oriented (anticipation) anxiety, such as generalized anxiety disorder (e.g., Luyten, Vansteenwegen, van Kuyck, & Nuttin, 2011). In addition, context conditioning might also explain agoraphobic avoidance often observed in individuals with panic disorder (e.g., Craske, Glover, & DeCola, 1995; Gorman, Kent, Sullivan, & Coplan, 2000). In particular, uncued panic attacks might serve as USs that are not predicted by discrete cues but merely occur in a particular context (e.g., a supermarket). As a consequence, people might come to avoid the entire context and related contexts (e.g., crowded places), engaging in generalized avoidance. Therefore, one component in the treatment of panic disorder is to identify discrete exteroceptive and interoceptive cues that are predictive for panic attacks (Craske & Barlow, 2008). Fonteyne, Vervliet, Baeyens, and Vansteenwegen (2009) provided experimental evidence for the importance of this treatment component. They used a context conditioning procedure in which both groups received predictable shocks (i.e., paired with a discrete CS) in context A and shocks that were presented unpaired with a discrete CS in context B. As predicted, higher contextual fear was observed in context B as compared to context A. To investigate whether increasing the predictability of the US would reduce contextual fear, in

[5] For a more elaborate review of complex conditioning procedures, we refer the interested reader to Boddez, Baeyens, Hermans, and Beckers (2014).

a subsequent phase the shocks were signaled by a novel discrete CS in context B. Results showed that this intervention led to a reduction of contextual fear and therefore confirm that increasing predictability can decrease contextual fear.

A second set of complex conditioning procedures relates to discriminating between danger and safety. In real life, the ability to discriminate between stimuli that signal danger and stimuli that indicate the absence of danger can be considered highly adaptive. For instance, the symptoms of a panic attack might resemble a life-threatening heart attack or a stroke, but are in fact innocuous (Haddad, Pritchett, Lissek, & Lau, 2012). Responding to a panic attack as if it is a heart attack or a stroke, leads to unnecessary escape and avoidance and an inefficient use of resources. The ability to discriminate between danger and safety cues can be modelled in a *differential inhibition* procedure, in which one stimulus (CS+) is paired with an aversive outcome, whereas another stimulus (CS-) predicts the absence of an aversive outcome. Using this procedure, elevated fear responding to safe stimuli (CS-) and impaired discrimination between danger and safety cues have been found in individuals with subclinical levels of anxiety (e.g., Ganzendam, Kamphuis, & Kindt, 2013; Haddad et al., 2012) and in patients with clinical anxiety (Duits et al., 2015; Lenaert, Boddez, Vervliet, Schruers, & Hermans, 2015; Lissek et al., 2005; Lissek et al., 2009). This suggests that the differential inhibition procedure taps into a process that is relevant to anxiety disorders and therefore has construct validity. Nonetheless, it should also be mentioned that some fear conditioning studies failed to find an effect of trait anxiety or observed an effect in only one of the outcome measures but not in the other(s) (e.g., Kindt & Soeter, 2014; Torrents-Rodas et al., 2013).

Other procedures have been used to examine impairments in safety learning. One of them is the *conditioned inhibition procedure*, a procedure known from the animal literature. In intermixed trials, one stimulus A is paired with the US, but when this stimulus is presented together with another stimulus B the US is omitted. This procedure corresponds to the use of safety signals in clinical practice. For example, an individual with driving phobia might be fearful of driving on a highway (A), but feel safe when accompanied with a fellow passenger (AB). Here as well, it has been found that high anxious individuals show higher fear responding to the safe AB stimulus compound than low anxious individuals (Chan & Lovibond, 1996; Grillon & Ameli, 2001).

A third defining feature of anxiety disorders is *stimulus generalization*. Boddez, Bennett, van Esch, and Beckers (2017) proposed to speak of generalization when a stimulus elicits a response due to a learning experience in which that stimulus as such was not featured. As an

example, consider a child experiencing painful skin burns upon touching the stove in one's grandparents' place. Behaving cautious around not just this but any stove will prevent the child from acquiring additional skin burns. In anxiety disorders, however, fear responding typically spreads to a range of stimuli and situations that are not dangerous (Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015; Hermans, Baeyens, & Vervliet, 2013). An individual suffering from dog phobia, for instance, will typically not only react with intense fear to the specific dog involved in the biting incident (who has proven to be dangerous), but to each and every dog.

In *perceptual stimulus generalization*, the fear responding is elicited by stimuli that share perceptual similarity with the original CS+ (Kalish, 1969; Lissek et al., 2010). The example of the person suffering from dog phobia discussed above can be considered an example of perceptual generalization. In a fear conditioning paradigm, Lissek et al. (2008) investigated perceptual generalization by presenting small and large sized circles as CS+ and CS- respectively, counterbalanced across participants. Subsequently, perceptual generalization of fear responding was tested by presenting participants with circles that ranged in size between the CS+ and CS- (i.e., generalization stimuli). This generated a generalization gradient, with stronger fear responses to stimuli that resemble the CS+ and decreasing responding with decreasing similarity. Using this procedure, relatively stronger responding to the generalization stimuli (GSs) has been observed in patients with panic disorder (Lissek et al., 2010), generalized anxiety disorder (Lissek, et al., 2014) and posttraumatic stress disorder (Lissek & Grillon, 2012) compared to healthy controls. Taken together, this suggests that the generalization procedure seizes a mechanisms that is relevant to anxiety disorders and therefore is construct valid.

It is important to note that perceptual generalization can also involve interoceptive CSs (Dymond et al., 2015; Lissek et al., 2010; Schroijen & Pappens, 2015). In a patient suffering from panic disorder, fear responding might be elicited not only by the interoceptive symptoms that directly preceded and accompanied the first panic attack (e.g., tight chest), but also by other physical symptoms (e.g., full stomach after overeating).

In addition to these different forms of perceptual generalization, fear responses can also generalize across stimuli that diverge greatly in perceptual features, based on nonperceptual grounds such as categories and conceptual knowledge (e.g. Dunsmoor & Murphy, 2015). Alex not only avoided going to the class of the specific professor that was involved in the incident, he also skipped other classes and started to avoid all kinds of (social) situations including going

to activities of the student club and meeting new people. Although these situations might clearly differ from the original learning experience with regard to their physical characteristics, they can be considered part of an idiosyncratic category of situations in which Alex considers himself at risk of saying something stupid and being rejected by others. This type of generalization is called *nonperceptual-based generalization* (Dymond et al., 2015).

As an experimental illustration, Dunsmoor, Martin, and LaBar (2012) presented participants with a heterogeneous collection of images of animals and tools. Images of one of these categories (e.g., animals) served as the CS+ category and were paired with an electric shock, the other stimulus category (e.g., tools) was never reinforced (i.e., CS- category). Subsequently, they tested for the generalization of fear to new stimuli stemming from the CS+ and CS- category. Results indicated higher fear responding, as measured by shock expectancy and skin-conductance response, to stimuli stemming from the CS+ category compared to stimuli from the CS- category. These results suggest that fear can generalize based on categorical and conceptual knowledge. However, this paradigm using stimuli from preexisting categories does not allow to rule out the potential influence of perceptual similarity on the generalization of fear. It can indeed be argued that all stimuli of one category (e.g., animals) share more perceptual similarity with each other than with stimuli from the other category (e.g., tools).

To completely rule out the potential influence of perceptual overlap, research has been conducted using de novo categories by inducing concept-like relations between arbitrary stimuli in an experimental way. Vervoort, Vervliet, Bennett, and Baeyens (2014) investigated the generalization of fear acquisition within novel arbitrary categories by first creating two four-member stimulus equivalence categories (i.e., A1-B1-C1-D1 and A2-B2-C2-D2). This was done using a matching-to-sample task. Stimuli were arbitrary line drawings. In the matching-to-sample task, a sample stimulus was presented together with two comparison figures and participants were instructed to choose the comparison figure that matched the sample stimulus. After every trial participants received feedback on whether or not their response was correct. Next, one member of the first category (B1) was presented repeatedly with an electric shock, whereas the member of the second category (B2) was never paired with shock. In a subsequent test for generalization in which C1, D1, C2 and D2 were presented, it was found that conditional fear responses generalized to other members of the arbitrary category (i.e., C1, D1). These results confirm that fears can generalize across conceptually-related in addition to perceptually-related stimuli.

Finally, Boddez et al. (2012) used a blocking procedure to assess fear generalization. In the first stage of the experiment, a CS was paired with a US. In a subsequent phase, a second CS was presented in compound with the first CS. This compound was paired with the same US as in the first phase. The newly added second CS therefore did not provide any information about the onset of the US over and above the information provided by the first CS. Interestingly, high trait anxious participants showed higher fear responding when this second CS was presented by itself at test. Such deficit in blocking might explain why certain individuals are more prone to develop clinical anxiety. A soldier might, for example, have experienced a bomb attack preceded by different cues such as a screaming colleague warning for the attack, a sandy surface, and fire of the bombing raid. A deficit in blocking would imply that the soldier, after his mission, does not only experience anxiety when a colleague is warning for another bomb attack (i.e., the most informative cue), but also when he is invited to a barbecue (fire) or when visiting the beach (sand).

In summary, we discussed complex learning procedures that allow to mimic specific processes at play in anxiety disorders. As such, these procedures add to the construct validity of the fear conditioning approach.

Predictive validity. Arguments about theoretical processes notwithstanding, an important question remains whether the fear conditioning model allows for translation to real-life situations that pertain to clinical anxiety. This concerns the predictive validity of the fear conditioning model and will be discussed now. Two aspects of predictive validity can be distinguished (Scheveneels et al., 2016). A first aspect refers to whether environmental variables exert a similar influence in the fear conditioning model and in real-life situations. A second aspect concerns testing whether a factor at the level of the individual exerts such similar influence as well. Both aspects are about the question of whether a factor, be it an environmental intervention or at the level of the individual, moderates the relation between learning experiences and fear in the lab in the same way as the relation between aversive experiences and anxiety symptoms in real life.

Do interventions have a similar effect in the fear conditioning model and in real life? With regard to the first aspect of predictive validity, presenting the CS without US following fear conditioning is known to result in a decrease in fear responding (e.g., Hermans, Craske, Mineka, & Lovibond, 2006). This procedure is termed fear extinction. Interestingly, in real life a similar intervention also results in a decrease in fear responding, thus adding to the predictive

validity of the fear conditioning account. Indeed, exposure therapy or the repeated and systematic confrontation with the feared stimulus or situation without occurrence of the expected aversive outcome is the (psychological) treatment of choice in anxiety (e.g., Cusack et al., 2016; Öst, Havnen, Hansen, & Kvale, 2015, Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008). During treatment, Alex was exposed to those social situations in which he expected to be rejected by others, such as making telephone calls when his colleagues are in the same room, meeting new people at parties, returning things to shops, etc. After repeated confrontation with these situations, Alex reported to experience less anxiety.

In addition, experimental research has demonstrated that fear responding can return after (partial or complete) fear extinction (Rachman, 1989, Vervliet, Craske, & Hermans, 2013). In clinical practice, a return in fear responding is unfortunately not uncommon either. It is estimated that 19-62% of clients experience at least some return of fear after exposure-based treatment (Craske & Mystkowski, 2006). Interesting for our present purposes, the interventions that cause a return of fear after extinction in experimental conditioning studies correspond to pathways to return of fear in real life. This again vows for the predictive validity of the model (Vervliet et al., 2013). We will now discuss the most well-studied pathways to return of fear: spontaneous recovery, (context) renewal, and reinstatement.

In the laboratory return of fear can occur when a time interval is introduced after extinction. This phenomenon has been described by Pavlov (1927) as *spontaneous recovery* and has been established in multiple laboratory studies (e.g., Huff, Hernandez, Blanding, & LaBar, 2009; Norrholm et al., 2008). Spontaneous recovery in the lab corresponds to the clinical observation that due to the mere passage of time after exposure treatment a client can show a reappearance of fearful responding (e.g., Mystkowski, Craske, Echiverri, & Labus, 2006; Vasey, Harbaugh, Buffington, Jones, & Fazio, 2012).

A second manipulation that causes return of fear and has been investigated in fear conditioning research is a change in (background) context between extinction and a subsequent test phase. This is referred to as (contextual) *renewal* (e.g., Bouton, 2002; Effting & Kindt, 2007; Vervliet, Baeyens, Van den Bergh, & Hermans, 2013). Typically, if CS-US pairings during acquisition take place in context A and fear is extinguished in context B, return of fear responding is observed if the CS is presented in the acquisition context A (ABA renewal) or in a novel context C (ABC renewal) (Bouton & Bolles, 1979). In clinical practice, this corresponds to a relapse after successful treatment when the feared object or situation is encountered outside the treatment context. Mystkowski, Craske, and Echiverri (2002) investigated return of fear

after a context change in a sample of spider-anxious individuals. All participants received a 1-session exposure-based therapy in which they were exposed to a spider in one particular context. Fear responding to the spider was tested one week later in the treatment context as well as in a novel context. Significantly higher fear responding was observed in the novel context as compared to the treatment context. Importantly, also the therapist can be considered as a contextual factor: being confronted with the feared stimulus or situation in the absence of the therapist after (successful) treatment can result in a return of fear responding (Rodriguez, Craske, Mineka, & Hladek, 1999). Based on these findings, extinction in multiple contexts has shown to reduce renewal in the laboratory (e.g., Bandarian-Balooch, Neumann, & Boschen, 2012). Similarly, exposing a patient to the feared stimulus in multiple contexts during clinical treatment (e.g., in the treatment context, at home etc.) can attenuate return of fear responding (e.g., Olatunji, Tomarken, Wentworth, & Fritsche, 2017; Vansteenwegen, Vervliet, Hermans, Thewissen, & Eelen, 2007).

A third intervention to induce return of fear is *reinstatement*. This refers to a return in fear responding due to unsignaled US-presentations after extinction (Haaker, Golkar, Hermans, & Lonsdorf, 2014; Hermans et al., 2005). Reinstatement can be seen as the equivalent of a return of fear after unsignaled panic attacks or if the previously feared stimulus is encountered after a stressful event or in a distressing situation. After successful treatment, Alex experienced a short-term recurrence of social anxiety after someone commented him about his blushing face.

A similar argument can be made with regard to pharmacological interventions for anxiety. If the fear conditioning model has predictive validity, a pharmacological intervention that has shown to effectively reduce clinical anxiety, should exert a similar effect in the fear conditioning model. Amongst the leading pharmacological interventions for anxiety are benzodiazepines (e.g., Offidani, Guidi, Tomba, & Fava, 2013). In the fear conditioning model, there is evidence that contextual fear (but not CS-specific fear) can be reduced by benzodiazepines such as alprazolam (Baas et al., 2002; Grillon et al., 2006). These results suggest that more complex learning procedures such as context conditioning are more useful to test the clinical utility of anxiolytic interventions compared to the simple acquisition procedure.

Do factors at the level of the individual have a similar effect in the fear conditioning model and in real life? This aspect of predictive validity concerns testing whether a factor at the level of the individual moderates the relation between learning experiences and fear in the lab in the same way as the relation between learning experiences and anxiety in real life.

In a longitudinal design, Lenaert et al. (2014) tested whether generalization and discrimination learning in a human fear conditioning procedure could predict subclinical levels of anxiety at 6-months follow-up. A large sample of first-year students completed a differential inhibition procedure followed by a generalization test. US-expectancy ratings were used as the outcome measure. Lenaert et al. (2014) argue that first-year students are particularly interesting because the transition to university is accompanied by a set of real-life aversive experiences related to academics, finances, social interaction, and other issues. Crucially, students who showed deficiencies in discriminating between the CS+ and CS- in the differential inhibition procedure (i.e., impaired safety learning) reported higher levels of anxiety at 6-month follow-up. In addition, elevated responding to the generalization stimuli closer to the CS- predicted higher levels of anxiety at follow-up. These findings suggest that these complex conditioning procedures have predictive validity. More precisely, it seems that the way in which characteristics of the individual affect the effect of aversive experiences on fear expression is the same in the laboratory as in real life.

In other studies, soldiers and firemen were used as subjects. The logic underlying these studies was the same: to assess whether individuals react similarly to aversive learning experiences in the lab as they do in real life. Needless to say, soldiers and firemen are confronted with a plethora of aversive learning experiences, making this an interesting population. Sijbrandij, Engelhard, Lommen, Leer, and Baas (2013) found that impaired safety learning in the lab was associated with PTSD symptoms at 2 and 9 months post-deployment to Afghanistan. Similar findings were reported by Lommen, Engelhard, Sijbrandij, van den Hout, and Hermans (2013) and by Acheson et al. (2015). Guthrie and Bryant (2006), on their turn, found that deficits in extinction learning in the lab are a risk factor for PTSD symptoms after trauma exposure in firemen.

In conclusion, performance in the fear conditioning model allows to predict (sub)clinical levels of anxiety, pointing towards a key role for conditioning in (clinical) anxiety. Importantly, this paves the way for targeted prevention in individuals who are at relatively higher risk.

Conclusion

Fear conditioning procedures have been applied extensively as a model for the acquisition of (clinical) fears and anxiety. In this chapter, we described the fear conditioning model and evaluated its external validity based on three validity criteria: face validity, construct validity and predictive validity. The fear conditioning model shows sufficient face validity and allows

for further increasing of the similarities between the fear conditioning model and clinical anxiety by including technologies such as virtual reality. Some critiques have been formulated with regard to whether the (etiological) processes that underlie the fear conditioning model are the same as those at work in clinical anxiety (i.e., construct validity). In particular, the simple fear acquisition model, as proposed by Watson, might be insufficient to explain why some individuals develop an anxiety disorders and others do not. We discussed how modern learning approaches have addressed these criticisms by, amongst other things, taking into account contextual variables before, during or after the conditioning experience. Furthermore, the use of more complex conditioning procedures might add to the construct validity of the fear conditioning model by mimicking additional processes at play in anxiety disorders. In the section on predictive validity, we discussed that environmental and individual-level factors that decrease and increase fear after aversive experiences in the lab also do so in real life. In conclusion, the fear conditioning approach allows to investigate the acquisition of fear under highly controlled circumstances and makes it possible to identify the exact (learning) mechanisms involved in the etiology of anxiety disorders. This knowledge can provide meaningful directions in how to prevent and treat (clinical) anxiety.

References

Acheson, D. T., Geyer, M. A., Baker, D. G., Nievergelt, C. M., Yurgil, K., & Risbrough, V. B. (2015). Conditioned fear and extinction learning performance and its association with psychiatric symptoms in active duty Marines. *Psychoneuroendocrinology, 51*, 495-505.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, D. C.: American Psychiatric Association.

Arnoudova, 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. doi:10.3389/fpsyg.2013.00298

Baas, J. M., Grillon, C., Böcker, K. B., Brack, A. A., Morgan, C. A., Kenemans, L. J., & Verbaten, M. N. (2002). Benzodiazepines have no effect on fear-potentiated startle in humans. *Psychopharmacology, 161*, 233-247. doi:10.1007/s00213-002-1011-8

Baas, J. M., Nugent, M., Lissek, S., Pine, D. S., & Grillon, C. (2004). Fear conditioning in virtual reality contexts: A new tool for the study of anxiety. *Biological Psychiatry, 55*, 1056-1060. doi:10.1016/j.biopsych.2004.02.024

Bandarian-Balooch, S., Neumann, D. L., & Boschen, M. J. (2012). Extinction treatment in multiple contexts attenuates return ABC renewal in humans. *Behaviour Research and Therapy, 50*, 604-609. doi:10.1016/j.brat.2012.06.003

Bauer, E. P. (2015). Serotonin in fear conditioning processes. *Behavioural Brain Research, 277*, 68-77. doi:10.1016/j.bbr.2014.07.028

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

Blaya, C., Salum, G. A., Lima, M. S., Leistner-Segal, S., Manfro, G. G. (2007). Lack of association between the serotonin transporter promotor polymorphism (5-HTTLPR) and panic disorder: A systematic review and meta-analysis. *Behavioral and Brain Functions, 41*. doi:10.1186/1744-9081-3-41

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology, 42*, 1-15.

Boddez, Y., Baeyens, F., Hermans, D., & Beckers, T. (2014). A fear conditioning approach to

anxiety disorders: The added value of complex acquisition procedures. In P. Emmelkamp & T. Ehring (Eds.), *The Wiley Handbook of Anxiety Disorder* (pp. 85-103). New York, NY: Wiley-Blackwell.

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,* 201-206. doi:10.1016/j.jbtrp.2012.08.003

Boddez Y., Bennett M., van Esch S., Beckers T. (2017). Bending rules: The shape of the perceptual generalization gradient is sensitive to inference rules. *Cognition & Emotion, 31,* 1444-1452.

Boddez, Y., Vervliet, B., Baeyens, F., Lauwers, S., Hermans, D., & Beckers, T. (2011). Expectancy bias in a selective conditioning procedure: Trait anxiety increases the threat value of a blocked stimulus. *Journal of Behavior Therapy and Experimental Psychiatry, 43,* 832-837. doi:10.1016/j.jbtep.2011.11.005

Bouman, T. K., & van Hout, W. J. P. J. (2006). CS-exposure werkt bij emetofobie. *Gedragstherapie, 39,* 127-138.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52,* 976-986.

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10,* 445-466.

Bouton, M. E., Mineka, S., & Barlow, D. H. (2001). A modern learning theory perspective on the etiology of panic disorders. *Psychological Review, 108,* 4–32.

Cameron, G., Roche, B., Schlund, M. W., & Dymond, S. (2016). Learned, instructed and observed pathways to fear and avoidance. *Journal of Behavior Therapy and Experimental Psychiatry, 50,* 106-112. doi:10.1016/j.jbtep.2015.06.003

Carleton, R. N. (2016). Into the unknown: A review and synthesis of contemporary models involving uncertainty. *Journal of Anxiety Disorders, 39,* 30-43. doi:10.1016/j.anxdis.2016.02.007

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

Chin, B., Nelson, B. D., Jackson, F., & Hajcak, G. (2016). Intolerance of uncertainty and startle potentiation in relation to different threat reinforcement rates. *International Journal of Psychophysiology, 99,* 79-84. doi:10.1016/j.ijpsycho.2015.11.006

Craske, M. G., & Barlow, D. H. (2008). Panic disorder and agoraphobia. In D. H. Barlow

(Ed.), *Clinical Handbook of Psychological Disorders* (pp. 1-64). New York, NY: The Guilford Press.

Craske, M. G., Glover, D., & DeCola, J. (1995). Predicted versus unpredicted panic attacks: Acute versus general stress. *Journal of Abnormal Psychology, 104*, 214-223.

Craske, M. G., & Mystkowski, J. L. (2006). Exposure therapy and extinction: Clinical studies. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and learning: From basic processes to clinical implications* (pp. 217-233). Washington, DC: American Psychiatric Association.

Cusack, K., Jonas, D. E., Fomeris, C. A., Wines, C. Sonis, J., Middleton, J. C., …, Gaynes, B. N. (2016). Psychological treatments for adults with posttraumatic stress disorder: A systematic review and meta-analysis. *Clinical Psychology Review, 43*, 128-141. doi:10.1016/j.cpr.2015.10.003

Damsa, C., Kosel, M., & Moussally, J. (2009). Current status of brain imaging in anxiety disorders. *Current Opinion in Psychiatry, 22*, 96-110. doi:10.1097/YCO.0b013e328319bd10

Davey, G. C. (2017). A research pathway for experimental psychopathology: the role of external validity criteria. *Psychopathology Review, 4*, 129-140.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review, 20*, 631-642. doi:10.3758/s13423-013-0386-3

den Hollander, M., Meulders, A., Jakobs, M., & Vlaeyen, J. W. S. (2015). The effect of threat information on acquisition, extinction and reinstatement of experimentally conditioned fear of movement-related pain. *Pain Medicine, 16*, 2302-2315. doi:10.1111/pme.12839

Dick, D. M., Agrawal, A., Keller, M. C., Adkins, A., Aliev, F., Monroe, S., … Sher, K. J. (2015). Candidate gene-environment interaction research reflections and recommendations. *Perspectives on Psychological Science, 10*, 37-59. doi:10.1177/1745691614556682

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

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *The American Journal of Psychiatry, 168*, 1041-1049. doi:10.1176/appi.ajp.2011.11020191

Dunsmoor, J. E., Martin, A., & LaBar, K. S. (2012). Role of conceptual knowledge in learning and retention of conditionned fear. *Biological Psychology, 89*, 300-305. doi:10.1016/j.biopsycho.2011.11.002

Dunsmoor, J. E., & Murphy, G. L. (2015). Categories, concepts, and conditioning: How humans generalize fear. *Trends in Cognitive Sciences, 19*, 73-75. doi:10.1016/j.tics.2014.12.003

Dymond, Dunsmoor, Vervliet, Roche, & Hermans (2015). Fear generalization in humans: Systematic review and implications for anxiety disorder research. *Behavior Therapy, 46*, 561-582.

Effting, M., & Kindt, M. (2007). Contextual control of human fear associations in a renewal paradigm. *Behaviour Research and Therapy, 45*, 2002-2018. doi:10.1016/j.brat.2007.02.011

Engelhard, I. M., van den Hout, M. A., & McNally, R. J. (2008). Memory consistency for traumatic events in Dutch soldiers deployed to Iraq. *Memory, 16*, 3-9.

Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *The American Journal of Psychiatry, 164*, 1476-1488. doi:10.1176/appi.ajp.2007.07030504

Fonteyne, R., Vervliet, B., Hermans, D., Baeyens, F., & Vansteenwegen, D. (2009). Reducing chronic anxiety by making the threatening event predictable: An experimental approach. *Behaviour Research and Therapy, 47*, 830-839. doi:10.1016/j.brat.2009.06.011

Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.

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

Gentes, E. L., & Ruscio, A. M. (2011). A meta-analysis of the relation of intolerance of uncertainty to symptoms of generalized anxiety disorder, major depressive disorder, and obsessive-compulsive disorder. *Clinical Psychology Review, 31*, 923-933. doi:10.1016/j.cpr.2011.05.001

Gorman, J. M., Kent, J. M., Sullivan, G. M., & Coplan, J. D. (2000). Neuroanatomical hypothesis of panic disorder, revised. *American Journal of Psychiatry, 157*, 493-505.

Grillon, C. (2002). Startle reactivity and anxiety disorders: Aversive conditioning, context, and neurobiology. *Biological Psychiatry, 52*, 958-975.

Grillon, C., & Ameli, R. (2001). Conditioned inhibition of fear-potentiated startle and skin conductance in humans. *Psychophysiology, 38*, 807-815.

Grillon, C., & Baas, J. (2003). A review of the modulation of the startle reflex by affective states and its application in psychiatry. *Clinical Neurophysiology, 114*, 1557-1579.

Grillon, C., Baas, J., Lissek, S., Smith, K., & Milstein, J. (2004). Anxious responses to predictable and unpredictable aversive events. *Behavioral Neuroscience, 118*, 916-924. doi:10.1037/0735-7044.118.5.916

Grillon, C., Baas, J., Pine, D. S., Lissek, S., Lawley, M., Ellis, V., & Levine, J. (2006). The benzodiazepine alprazolam dissociates contextual fear from cued fear in humans as assessed by fear-potentiated startle. *Biological Psychiatry, 60*, 760-766. doi:10.1016/biopsych.2005.11.027

Grillon, C., & Davis, M. (1997). Fear-potentiated startle conditioning in humans: Explicit and contextual cue conditioning following paired versus unpaired training. *Psychophysiology, 34*, 451-458. doi:10.1111/j.1469-8986.1997.tb02389.x

Guthrie, R. M., & Bryant, R. A. (2006). Extinction learning before trauma and subsequent posttraumatic stress. *Psychosomatic medicine, 68*, 307-311.

Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014). A review on human reinstatement studies: An overview and methodological challenges. *Learning and Memory, 21*, 424-440. doi:10.1101/lm.036053.114

Haddad, A. D. M., Pritchett, D., Lissek, S., & Lau, J. Y. F. (2012). Trait anxiety and fear responses to safety cues: Stimulus generalization of sensitization? *Journal of Psychopathology and Behavioral Assessment, 34*, 323-331. doi:10.1007/s10862-012-9284-7

Hedge, C., Powell, G., Sumner, P., (2017). The reliability paradox: Why robust cognitive tasks do produce reliable individual differences. *Behavior Research Methods.* doi:10.3758/s13428-017-0935-1

Hermans, D., Baeyens, F., & Vervliet, B. (2013). Generalization of acquired emotional responses. In M. D. Robinson, E. R. Watkins, & E. Harmon-Jones (Eds.), *Handbook of cognition and emotion* (pp. 117-134). New York, NY: Guilford Press.

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

Hermans, D., Dirickx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behavior Research and Therapy, 43*, 533-551. doi:10.1016/j.brat.2004.03.013

Hodgson, R., & Rachman, S. (1974). Desynchrony in measures of fear. *Behaviour Research and Therapy, 12*, 319-326.

Huff, N. C., Hernandez, J. A., Blanding, N. Q., & LaBar, K. S. (2009). Delayed extinction attenuates conditioned fear renewal and spontaneous recovery in humans. *Behavioral Neuroscience, 123*, 834-843. doi:10.1037/a0016511

Ipser, J. C., Singh, L., Stein, D. J. (2013). Meta-analysis of functional brain imaging in specific phobia. *Psychiatry and Clinical Neurosciences, 67*, 311-322. doi:10.1111/pcn.12055

Joos, E., Vansteenwegen, D., & Hermans, D. (2012). Post-acquisition repetitive thought in fear conditioning: An experimental investigation of the effect of CS-US rehearsal. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 737-744. doi:10.1016/j.btep.2011.10.011

Joos, E., Vansteenwegen, D., Vervliet, B., & Hermans, D. (2013). Repeated activation of a CS-US-contingency results in sustained conditioned responding. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00305

Kalish, H. (1969). Stimulus generalization. In M. Marx (Ed.), *Learning: Processes* (pp. 205-297). Oxford, UK: Macmillan.

Kent, G. (1997). Dental phobias. In G. C. Davey (Ed.), *Phobias: A handbook of theory, Research and Treatment* (pp. 107-127). Chichester, England: Wiley.

Kent, J. M., & Rauch, S. L. (2003). Neurocircuitry of anxiety disorders. *Current Psychiatry Reports, 5*, 266-273.

Kindt, M., & Soeter, M. (2014). Fear inhibition in high trait anxiety. *PLoS ONE, 9*, e86462. doi:10.1371/journal.pone.0086462

Krypotos, A.-M., Arnaudova, I., Effting, M., Kindt, M., & Beckers, T. (2015). Effects of approach-avoidance training on the extinction and return of fear responses. *PLoS ONE, 10*, e0131581. doi:10.1371/journal.pone.0131581

Krypotos, A.-M., Effting, M., Arnaudova, I., Kindt, M., & Beckers, T. (2014). Avoided by association: Acquisition, extinction, and renewal of avoidance tendencies towards conditioned fear stimuli. *Clinical Psychological Science, 2*, 336-343. doi:10.1177/2167702613503139

Lang, P. J. (1979). A bio-informational theory of emotional imagery. *Psychophysiology, 16*, 495-512.

Leibold, N. K., Viechtbauer, W., Goossens, L., De Cort, K., Griez, E. J., Myin-Germeys, I., &

Steinbusch, H. W. M. (2013). Carbon dioxide inhalation as a human experimental model of panic: The relationship between emotions and cardiovascular physiology. *Biological Psychology, 94*, 331-340. doi:10.1016/j.biopsycho.2013.06.004

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

Lenaert, B., Boddez, Y., Vervliet, B., Schruers, K., & Hermans, D. (2015). Reduced autobiographical memory specificity is associated with impaired discrimination learning in anxiety disorder patients. *Frontiers in Psychology, 6*, 889. doi:10.3389/fpsyg.2015.00889

Lenaert B., Boddez Y., Vlaeyen J., & van Heugten C. (2018). Learning to feel tired: A learning trajectory towards chronic fatigue. *Behaviour Research and Therapy, 100*, 54-66

Leuchs, L., Schneider, M., Czisch, M., & Spoormaker, V. I. (2017). Neural correlated of pupil dilation during human fear learning. *NeuroImage, 147*, 186-197. doi:10.1016/j.neuroimage.2016.11.072

Lipp, O. V. (2006). Human fear learning: contemporary procedures and measurement. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and learning: From basic processes to clinical implications* (pp. 37-52). Washington, DC: American Psychological Association.

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

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

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biological Psychiatry, 75*, 909-956. doi:10.1016/j.biopsych.2013.07.025

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., &

Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behaviour Research and Therapy, 43*, 1391-1424.

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *American Journal of Psychiatry, 167*, 47-55. doi:10.1176/appi.ajp.2009.09030410

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

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*, 63-67.

Lonsdorf, T. B., & Baas, J. M P. (2015). Genetics in experimental psychopathology: From laboratory models to therapygenetics. Where do we go from here? *Psychopathology Review*.

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

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 & Biobehavioral Reviews*. doi:10.1016/j.neubiorev.2017.07.007

Lonsdorf, T. B., Rück, C., Bergström, J., Andersson, G., Öhman, A., Schalling, M., & Lindefors, N. (2009). The symptomatic profile of panic disorder is shaped by the 5-HTTLPR polymorphism. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 33*, 1479-1483. doi:10.1016/j.pnpbp.2009.08.004

Lonsdorf, T., Weike, A., Nikamo, P., Schalling, M., Hamm, A., & Öhman, A. (2009). Genetic gating of human fear learning and extinction: Possible implication for gene-environment interaction in anxiety disorder. *Psychological Science, 20*, 198-206. doi:10.1111/j.1467-9280.2009.02280.x

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. doi:10.1080/17470210701503229

Luyten, L., Vansteenwegen, D., van Kuyck, K., & Nuttin, B. (2011). Contextual conditioning in rats as an animal model for generalized anxiety disorder. *Cognitive, Affective, & Behavioral Neuroscience, 11*, 228-244. doi:10.3758/s13415-011-0021-6

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: The effect of nonreinforced pre-exposure to the conditional stimulus. *Journal of Comparative and Physiological Psychology, 52*, 415-419. doi:10.1037/h0046700

Mallan, K. M., Sax, J., Lipp, O. V. (2009). Verbal instruction abolishes fear conditioned to racial out-group faces. *Journal of Experimental Social Psychology, 45*, 1303-1307. doi:10.1016/j.jesp.2009.08.001

Manufò, M. R., Freimer, N. B., Ng, W., Ophoff, R., Veijola, J., Miettunen, J., … Flint, J. (2009). 5-HTTLPR genotype and anxiety-related personality traits: A meta-analysis and new data. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics, 150B*, 271-281. doi:10.1002/ajmg.b.30808

Maren, S. (2008). Pavlovian fear conditioning as a behavioral assay for hippocampus and amygdala function: caution and caveats. *European Journal of Neuroscience, 28*, 1661-1666. doi:10.1111/j.1460-9568.2008.06485.x

McEvoy, P. M., & Mahoney, A. E. J. (2012). To be sure, to be sure: Intolerance of uncertainty mediates symptoms of various anxiety disorders and depression. *Behvior Therapy, 43*, 533-545. doi:10.1016/j.beth.2011.02.007

Merckelbach, H., van den Hout, M. A., Hoekstra, R., & de Ruiter, C. (1989). Conditioning experiences and phobias. *Behaviour Research and Therapy, 27*, 657-662. doi:10.1016/0005-7967(89)90149-6

Meulders, A., Mampaey, J., Boddez, Y., Blanco, F., Vansteenwegen, D., & Baeyens, F. (2013). Offset-control attenuates context conditioning induced by US-unpredictability in a human conditioned suppression paradigm. *Psychologica Belgica, 53*, 39-56. doi:10.5334/pb-53-1-39

Mineka, S., Cook, M., & Miller, S. (1984). Fear conditioned with escapable and inescapable shock: Effects of a feedback stimulus. *Journal of Experimental Psychology: Animal Behavior Processes, 10*, 307-323.

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology

of anxiety disorders. *American Psychologist, 61*, 10-26. doi:10.1037/0003-066X.61.1.10

Miu, A. C., Vulturar, R., Chis, A., Ungureanu, L., & Gross, J. J. (2013). Reappraisal as a mediator in the link between 5-HTTLPR and social anxiety symptoms. *Emotion, 13*, 1012-1022. doi:10.1037/a0033383

Mystkowski, J. L., Craske, M. G., & Echiverri, A. M. (2002). Treatment context and return of fear in spider phobia. *Behavior Therapy, 33*, 399-416.

Mystowski, J. L., Craske, M. G., Echiverri, A. M., & Labus, J. S. (2006). Mental reinstatement of context and return of fear in spider-fearful participants. *Behavior Therapy, 37*, 49-60. doi:10.1016/j.beth.2005.04.001

Nelson, B. D., Weinberg, A., Pawluk, J., Gawlowska, M., & Proudfit, G. H. (2015). An event-related potential investigation of fear generalization and intolerance of uncertainty. *Behavior Therapy, 46*, 661-670. doi:10.1016/j.beth.2014.09.010

Norrholm, S. D., Vervliet, B., Jovanovic, T., Boshoven, W., Myers, K. M., Davis, M., …, & Duncan, E. J. (2008). Timing of extinction relative to acquisition: A parametric analysis of fear extinction in humans. *Behavioral Neuroscience, 122*, 1016-1030. doi:10.1037/a0012604

Offidani, E., Guidi, J., Tomba, E., & Fava, G. A. (2013). Efficacy and tolerability of benzodiazepines versus antidepressants in anxiety disorders: A systematic review and meta-analysis. *Psychotherapy and Psychosomatics, 82*, 355-362. doi:10.1159/000353198

Olatunji, B. O., Tomarken, A., Wentworth, B., & Fritzsche, L. (2017). Effects of exposure in single and multiple contexts on fear renewal: The moderating role of threat-specific and nonspecific emotionality. *Journal of Behavior Therapy and Experimental Psychiatry, 54*, 270-277. doi:10.1016/j.jbtep.2016.09.004

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fear by observing other: The neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience, 2*, 3-11. doi:10.1093/scan/nsm005

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychological Science, 15*, 822-828. doi:10.1111/j.0956-7976.2004.00762.x

Öst, L.- G., Havnen, A., Hansen, B., & Kvale, G. (2015). Cognitive behavioral treatment of

obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993-2014. *Clinical Psychology Review, 40*, 156-169. doi:10.1016/j.cpr.2015.06.003

Öst, L.- G., & Hugdahl, K. (1981). Acquistion of phobias and anxiety response patterns in clinical patients. *Behaviour Research and Therapy, 19*, 439-447.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex.* London: Oxford University Press.

Phelps, E. A., Connor, K. J. O., Gatenby, J. C., Gore, J. C., & Davis, M. (2001). Activation of the left to a cognitive representation of fear. *Nature Neuroscience, 4*, 437-441. doi:10.1038/86110

Pittig, A., Brand, M., Pawlikowski, M., & Alpers, G. W. (2014). The cost of fear: Avoidant decision making in a spider gambling task. *Journal of Anxiety Disorders, 28*, 326-334. doi:10.1016/j.anxdis.2014.03.001

Poulton, R., & Menzies, R. (2002). Non-associative fear acquisition: A review of the evidence from retrospective and longitudinal research. *Behaviour Research and Therapy, 40*, 127-149.

Rachman, S. J. (1977). The conditioning theory of fear acquisition: A critical examination. *Behaviour Research and Therapy, 15*, 375-387. doi:10.1016/0005-7967(77)90041-9

Rachman, S. J. (1989). The return of fear: Review and prospect. *Clinical Psychology Review, 9*, 147-168. doi:10.1016/0272-7358(89)90025-1

Rau, V., DeCola, J. P., & Fanselow, M. S. (2005). Stress-induced enhancement of fear learning: An animal model of posttraumatic stress disorder. *Neuroscience and Biobehavioral Reviews, 29*, 1207-1223. doi:10.1016/j.neubiorev.2005.04.010

Rescorla, R. A. (1974). Effect of inflation of the unconditioned stimulus value following conditioning. *Journal of Comparative and Physiological Psychology, 86*, 101-106.

Rinck, M., & Becker, E. S. (2006). Spider fearful individuals attend to threat, then quickly avoid it: Evidence from eye movements. *Journal of Abnormal Psychology, 115*, 231-238. doi:10.1037/0021-843X.115.2.231

Rodriguez, B. I., Craske, M. G., Mineka, S., & Hladek, D. (1999). Context-specificity of relapse: Effects of therapist and environmental context on return of fear. *Behaviour Research and Therapy, 37*, 845-862.

Scheveneels, S., Boddez, Y., Vervliet, B., & Hermans, D. (2016). The validity of laboratory-

based treatment research: Bridging the gap between fear extinction and exposure treatment. *Behaviour Research and Therapy, 86,* 87-94. doi:10.1016/j.brat.2016.08.015

Schroijen, M., & Pappens, M. (2015). Generalization of fear to respiratory sensations. *Behavior Therapy, 46,* 611-626. doi:10.1016/j.beth.2015.05.004

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

Shin, L. M., & Liberzon, I. (2010). The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology Reviews, 35,* 169-191. doi:10.1038/npp.2009.83

Sijbrandij, M., Engelhard, I. M., Lommen, M. J., Leer, A., & Baas, J. M. (2013). Impaired fear inhibition learning predicts the persistence of symptoms of posttraumatic stress disorder (PTSD). *Journal of psychiatric research, 47,* 1991-1997.

Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nature Reviews. Genetics, 3,* 391-397. doi:10.1038/nrg796

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. doi:10.1016/j.biopsycho.2012.10.006

Van Gucht, D., Vansteenwegen, D., Van den Bergh, O., & Beckers, T. (2008). Conditioned craving cues elicit an automatic approach tendency. *Behaviour Research and Therapy, 46,* 1160-1169.

van Meurs, B., Wiggert, N., Wicker, I., & Lissek, S. (2014). Maladaptive behavioral consequences of conditioned fear generalization: A pronounced, yet sparsely studied feature of anxiety pathology. *Behaviour Research and Therapy, 57,* 29-37. doi:10.1016/j.brat.2014.03.009

Vansteenwegen, D., Vervliet, B., Hermans, D., Thewissen, R., & Eelen, P. (2007). Verbal, behavioural and physiological assessment of the generalization of exposure-based fear reduction in a spider-anxious population. *Behavioural Research and Therapy, 45,* 291-300. doi:10.1016/j.brat.2006.03.008

Vasey, M. W., Harbaugh, C. N., Buffington, A. G., Jones, C. R., & Fazio, R. H. (2012).

Predicting return of fear following exposure therapy with an implicit measure of attitudes. *Behaviour Research and Therapy, 50,* 767-774. doi:10.1016/j.brat.2012.08.007

Vervliet, B., Baeyens, F., Van den Bergh, O., & Hermans, D. (2013). Extinction, generalization, and return of fear: A critical review of renewal research in humans. *Biological Psychology, 92,* 51-58. doi:10.1016/j.biopsycho.2012.01.006

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: State of the art. *Annual Review of Clinical Psychology, 9,* 215-248. doi:10.1146/annurev-clinpsy-050212-185542

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

Vervliet, B., & Raes, F. (2013). Criteria of validity in experimental psychopathology: Application to models of anxiety and depression. *Psychological Medicine, 43,* 2241-2244. doi:10.1017/S0033291712002267

Waters, A., LeBeau, R., & Craske, M. (2017). Experimental psychopathology and clinical psychology: An integrative model. *Psychopathology Review, 4,* 112-128. doi:10.5127/pr.038015

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

Wendt, J., Neubert, J., Lindner, K., Ernst, F. D., Homuth, G., Weike, A. I., Hamm, A. O. (2014). Genetic influences on the acquisition and inhibition of fear. *International Journal of Psychophysiology, 98,* 499-505. doi:10.1016/j.ijpsycho.2014.10.007

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

Chapter 2

One for all:

The effect of extinction stimulus typicality on return of fear

Based on:

Scheveneels, S., Boddez, Y., Bennett, M., & Hermans, D. (2017). One for all: The effect of extinction stimulus typicality on return of fear. *Journal of Behavior Therapy and Experimental Psychiatry, 57,* 37-44. doi:10.1016/j.jbtep.2017.03.002

Abstract

Background and Objectives. During exposure therapy, patients are encouraged to approach the feared stimulus, so they can experience that this stimulus is not followed by the anticipated aversive outcome. However, patients might treat the absence of the aversive outcome as an 'exception to the rule'. This could hamper the generalization of fear reduction when the patient is confronted with similar stimuli not used in therapy. We examined the effect of providing information about the typicality of the extinction stimulus on the generalization of extinction to a new but similar stimulus.

Methods. In a differential fear conditioning procedure, an animal-like figure was paired with a brief electric shock to the wrist. In a subsequent extinction phase, a different but perceptually similar animal-like figure was presented without the shock. Before testing the generalization of extinction with a third animal-like figure, participants were either instructed that the extinction stimulus was a typical or an atypical member of the animal family.

Results. The typicality instruction effectively impacted the generalization of extinction; the third animal-like figure elicited lower shock expectancies in the typical relative to the atypical group.

Limitations. Skin conductance data mirrored these results, but did not reach significance.

Conclusion. These findings suggest that verbal information about stimulus typicality can be a promising adjunctive to standard exposure treatments.

Keywords: Return of fear; Stimulus typicality; Fear conditioning; Fear extinction; Exposure therapy

One for all: The effect of the typicality of an extinction stimulus on return of fear

Exposure therapy involves the repeated confrontation with a fear-provoking stimulus and is the treatment of choice for anxiety disorders (e.g., Norton & Price, 2007; Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008). Nevertheless, some clients experience a re-emergence of fear symptoms after completing treatment (Craske & Mystkowski, 2006). One pathway for return of fear is the limited generalization of treatment effects when the client is confronted with a new stimulus from the feared category (Boddez et al., 2012; Rowe & Craske, 1998; Vervliet, Vansteenwegen, Baeyens, Hermans, & Eelen, 2005). For instance, a client suffering from dog phobia who has been successfully exposed to a German shepherd might experience a return of fear upon seeing an Airedale terrier. An important goal for clinical and translational research is, therefore, to enhance the generalization of exposure treatment to other exemplars of the feared category.

Extinction provides an elegant laboratory paradigm to tackle this type of question. In extinction, either a conditioned stimulus (CS; e.g., a geometrical figure) that was initially paired with an aversive unconditioned stimulus (US; e.g., electrical shock), or a stimulus that resembles this CS (i.e., a generalization stimulus; GS), is presented without the US (Hermans, Craske, Mineka, & Lovibond, 2006; Scheveneels, Boddez, Vervliet, & Hermans, 2016). This results in a decrease in the previously acquired (fear) responses. However, just like after initially successful exposure therapy, fear can return due to a variety of manipulations including confrontation with a stimulus different from the one used during extinction training (Vervliet, Craske, & Hermans, 2013; Vervoort, Vervliet, Bennett, & Baeyens, 2014). For example, in a study of Barry and colleagues (Barry, Griffith, Vervliet, & Hermans, 2016), a CS was paired with a US in the acquisition phase, so that it came to elicit fear responding. In the subsequent extinction phase, a perceptually similar GS was presented without the US. This GS elicited a high amount of fear responding at the beginning of the extinction phase, which, as expected, gradually decreased throughout the extinction phase. However, Barry et al. demonstrated that this successful extinction learning did not generalize to other stimuli: the presentation of yet another GS elicited fear responding again.

Existing extinction research already suggests one potential strategy to attenuate such return of fear after stimulus change in situations where fear is acquired via Pavlovian conditioning[1]: exposure to the stimulus to which fear was originally established (i.e., the CS)

[1] We refer the interested reader to McNally (2016) for a discussion of non-Pavlovian pathways to fear acquisition.

eliminates fear responding to other exemplars more effectively than exposure to a GS (Boddez et al., 2012; Dubin & Levis, 1975; Vervliet et al., 2005; Vervliet, Vansteenwegen, & Eelen, 2006). However, in clinical practice this original acquisition stimulus cannot always be identified or used (e.g., because of ethical considerations) and there is often no other option but to use a GS that more or less resembles the CS. In a study with spider phobics, Rowe and Craske (1998) therefore tested whether exposure to not just one but to multiple GSs (i.c., spiders) would enhance generalization of the observed reductions in fear responding to a new GS. As this did not turn out to be the case, we here propose another strategy to overcome this lack of transfer.

More precisely, we examined the effect of providing information about the typicality of the extinction stimulus on the generalization of extinction to a new GS. Indeed, principles of category-based induction suggest that properties of exemplars that are more representative or typical of the overall category are more likely to transfer to other category exemplars (i.e., GSs; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990). If a certain property holds for a typical bird (e.g., a sparrow), for example, then this property will be judged more likely to hold for other birds than if it holds for an atypical bird (e.g., a penguin). Recently, Dunsmoor and Murphy (2014) tested this in a conditioning procedure and showed that fear acquisition with a typical exemplar generalizes more broadly than fear acquisition with an atypical exemplar. We investigated whether enhancing the perceived categorical typicality of the extinction stimulus enhances the transfer of the reduction in fear responses to other exemplars.

It is of note that the dominant theory about extinction learning postulates that return of fear after extinction can be understood by assuming that subjects treat the omission of the US during extinction as a mere 'exception to the rule' that the US generally does follow the CS (Bouton, 2002; p. 982). If a GS is extinguished, for example, people might attribute the absence of the US to the fact that a stimulus other than the original CS was presented. From this perspective as well, informing people that the GS used during extinction is a typical exemplar might be a potentially promising strategy to violate this status of extinction learning as an 'exception to the rule'.

In the present study, we used a differential fear conditioning procedure in which a stimulus (CS) was paired with an electric shock (US). In a subsequent extinction phase, a perceptually similar GS from the same category-type (GS1) was presented without shock, after which another exemplar or GS from this category-type (GS2) was used to test the generalization of extinction. It was predicted that participants who received instructions about the extinction stimulus being a typical exemplar (i.e., the typical group) would show better generalization of

extinction learning to the test stimulus compared to participants who were instructed that the extinction stimulus is an atypical exemplar (i.e., the atypical group). US-expectancy ratings and skin conductance were measured as indices of fear.

Method

Participants

Sixty-nine undergraduate students at the University of Leuven (M_{age} = 21.30; *SD* = 4.18; 50 females) participated in exchange for course credits or payment (€ 8.00). This sample size was chosen in order to exceed the sample size that is typically reported in related fear conditioning studies (e.g., Barry et al., 2016; Boddez, Bennett, van Esch, & Beckers, 2016; Dunsmoor & Murphy, 2014). Participants were allocated in an alternating manner to the typical (*n* = 35) or atypical (*n* = 34) group: the first participant was assigned to the typical group, the second participant to the atypical group, the third participant again to the typical group, etc. All participants gave informed consent before starting the experiment and were aware that they could withdraw at any time. The standing ethical committee of the Faculty of Psychology and Educational Sciences approved the study.

Apparatus

Conditioned stimuli. Two separate families of artificial animal-like objects, known as 'Fribbles', served as experimental (CS+ family) and control (CS- family) stimuli (Barry, Griffith, De Rossi, & Hermans, 2014). Whether a particular family of Fribbles served as experimental or control stimuli was counterbalanced across participants. Six different Fribbles were used; one exemplar per family in each experimental phase (see Figure 1). All exemplars within a family had the same central body part but differed across phases with regard to their peripheral features (i.e., legs, head and tail). In particular, GS2 (used during test) shared two features with GS1 (used during extinction) and two features with the CS (used during acquisition; Barry et al., 2016). This overlap between exemplars in acquisition, extinction and test was analogous for stimuli in the CS+ and in the CS- family. With regard to notation, we use '+' and '-' to denote whether or not a US followed the CS in acquisition. In the extinction and test phase, we use CS+_GS1 and CS+_GS2 to indicate GSs from the CS+ family, and CS-_GS1 and CS-_GS2 to indicate GSs from the CS- family. Fribbles were approximately 8.45 cm

wide and 6.35 cm high and were presented against a black background on a 19 inch Dell monitor (type P1911, resolution: 1440 x 900 at 60 Hz).

Unconditioned stimulus. The US was a 2 ms electrocutaneous stimulus administered to the participant's right wrist by a Digitimer DS7A constant current stimulator (Hertfordshire, UK). Electrical stimulation was delivered via a pair of V91-01 8-mm reusable Bilaney Ag/AgCl electrodes filled with K-Y Jelly.

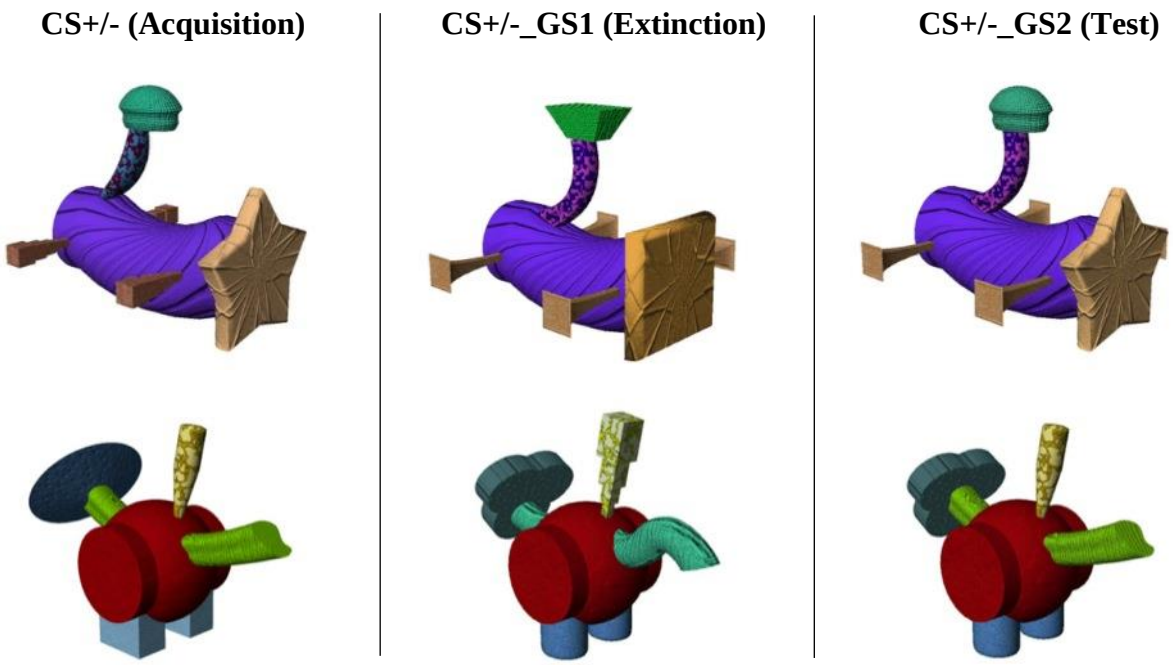

Figure 1. Six exemplars of two families of Fribbles were used as experimental and control stimuli (counterbalanced) in acquisition, extinction and test. During acquisition '+' refers to the administration of the US and '-' to the absence of the US. During the other phases no USs were administered, here CS+_GS1 and CS+_GS2 refer to GSs from the CS+ family (experimental stimuli) and CS-_GS1 and CS-_GS2 refer to GSs from the CS- family (control stimuli).

US-expectancy ratings. Participants rated their US-expectancy on a trial-by-trial basis on an 11-point scale, ranging from 0 = "certainly no shock" to 10 = "certainly shock". The scale was presented 200 ms after stimulus onset and remained onscreen for maximum 7 s. Responses

were made via a mouse-click and the scale disappeared once participants clicked a position on the scale.

Skin conductance response (SCR). Electrodermal responding was recorded by a Coulbourn LabLinc V Isolated Skin Conductance coupler (model V71-23, manufactured by Coulbourn Instruments, Allentown, PA) using the exosomatic method with alternating current excitation. The coupler applied a constant voltage of 0.5 Volts through a pair of disposable Biopac EL 507 electrodes (contact area = 95 mm^2) filled with isotonic paste. The electrodes were attached to the hypothenar site of the palm of the non-dominant hand. The analog signal was digitized at 10 Hz by a NI PCI-3221 data acquisition card (National Instruments Corporation, Austin, Texas). The resulting skin conductance data was analyzed offline with the Psychophysiological Analysis (PSPHA) software package (de Clercq, Verschuere, de Vlieger, & Crombez, 2006). SCR amplitudes (µSiemens) were calculated per trial by subtracting the mean value recorded in the 2 s immediately preceding stimulus onset (i.e., baseline period) from the maximum value recorded within 0-7.5 s following stimulus onset (Pineles, Orr, & Orr, 2009). Skin conductance data recorded during the actual presentation of the US (7.5-8 s) was not included in the analyses to avoid contamination and artefacts caused by the electrical stimulation. Difference scores were range-corrected with the participant's largest unconditioned response across shock trials (Lykken & Venables, 1971) and subsequently square root transformed to normalize the data (Dawson, Schell, & Fillion, 2000). Negative difference scores were recoded to zero.

Procedure

After participants gave informed consent, the electrodes were attached. Starting at 1.00 mA and moving upwards in steps of 1-2.00 mA, the US was calibrated to a level that was 'definitely uncomfortable, but not painful' (see Fonteyne, Vervliet, Hermans, Baeyens, & Vansteenwegen, 2010). In this study the mean intensity of the US was 26.00 mA (*SD* = 15.10). Before the experiment started, participants were given the following instructions:

"During this experiment you will be presented with 'Fribbles'. There are two families of Fribbles: the red ones and the purple ones. Some of the Fribbles are 'bad'. These bad Fribbles will bite you (i.e., they will be followed by an electric shock). Other Fribbles are 'good' Fribbles. These good Fribbles will not bite you, thus are not followed by an electric shock. It is your task to predict whether a particular Fribble will bite you or not."

Participants were then instructed about the rating scale and practiced using the scale by filling it in three times without presentation of a stimulus. Feedback on whether participants filled in the scale correctly was provided after each practice trial.

An overview of the experimental phases is given in Table 1. The experiment started with the pre-exposure phase that consisted of two non-reinforced presentations of each CS (i.e., CS+ and CS-) to reduce the orienting responding in the skin conductance measure. During acquisition each CS was presented six times: the CS+ was always paired with the US, whereas the CS- was never paired with the US. Extinction training followed immediately after acquisition, without participants being informed that a new experimental phase started. During extinction, each of the two GS1s (i.e., CS+_GS1 and CS-_GS1) was presented twelve times without the US. After extinction training, participants received information about the typicality of the extinction stimulus, depending on the group they were assigned to. Participants in the atypical group received the following instructions about the CS+_GS1 presented in extinction:

"Below you find additional information that will help you in predicting whether a Fribble will bite you or not. The red/purple Fribble you got to see recently was a special exemplar of the red/purple Fribbles. This Fribble is NOT characteristic for all red/purple Fribbles."

Participants assigned to the typical group were given the following instructions about the extinction CS+_GS1:

"Below you find additional information that will help you in predicting whether a Fribble will bite you or not. The red/purple Fribble you got to see recently was a typical exemplar of the red/purple Fribbles. This Fribble is characteristic for all red/purple Fribbles."

After this instruction, participants were tested with two new GSs (i.e., CS+_GS2 and CS-_GS2), each presented four times.

Table 1

Overview of the experimental phases

Pre-exposure	Acquisition	Extinction	Between-group manipulation	Test
CS+ (2)	CS+ (6)	CS+_GS1 (12)	Typicality instruction	CS+_GS2 (4)
CS- (2)	CS- (6)	CS-_GS1 (12)		CS-_GS2 (4)

Note. CSs and GSs are Fribbles (counterbalanced). The number of trials is indicated between parentheses.

Each trial started with a 2 s blank screen before stimulus onset for measuring baseline skin conductance. Fribbles remained onscreen for 8 s. On US-present trials, the shock was delivered at 7.5 s after stimulus onset. The inter-trial interval varied between 7 s and 9 s (average of 8 s). In all experimental phases stimuli were presented in semi-randomized order with no more than two consecutive trials with the same stimulus. Half of the participants in each group were presented with the CS+/CS+_GSs first in all phases, the other half with the CS-/CS-_GSs. This was done to control for order-effects: unreinforced CS+/CS+_GS presentations can have an impact on subsequent CS-/CS-_GS ratings (i.e., elevated ratings), leading to less discrimination between the CS/CS+_GS and CS-/CS-_GS (e.g., Lovibond, 2003; Vervliet et al., 2005).

Results

US-expectancy ratings

Figure 2 indicates the mean US-expectancy ratings in response to CSs and GSs, for both groups. Displayed are the first and last acquisition trials, the first and last extinction trials, and the four test trials.

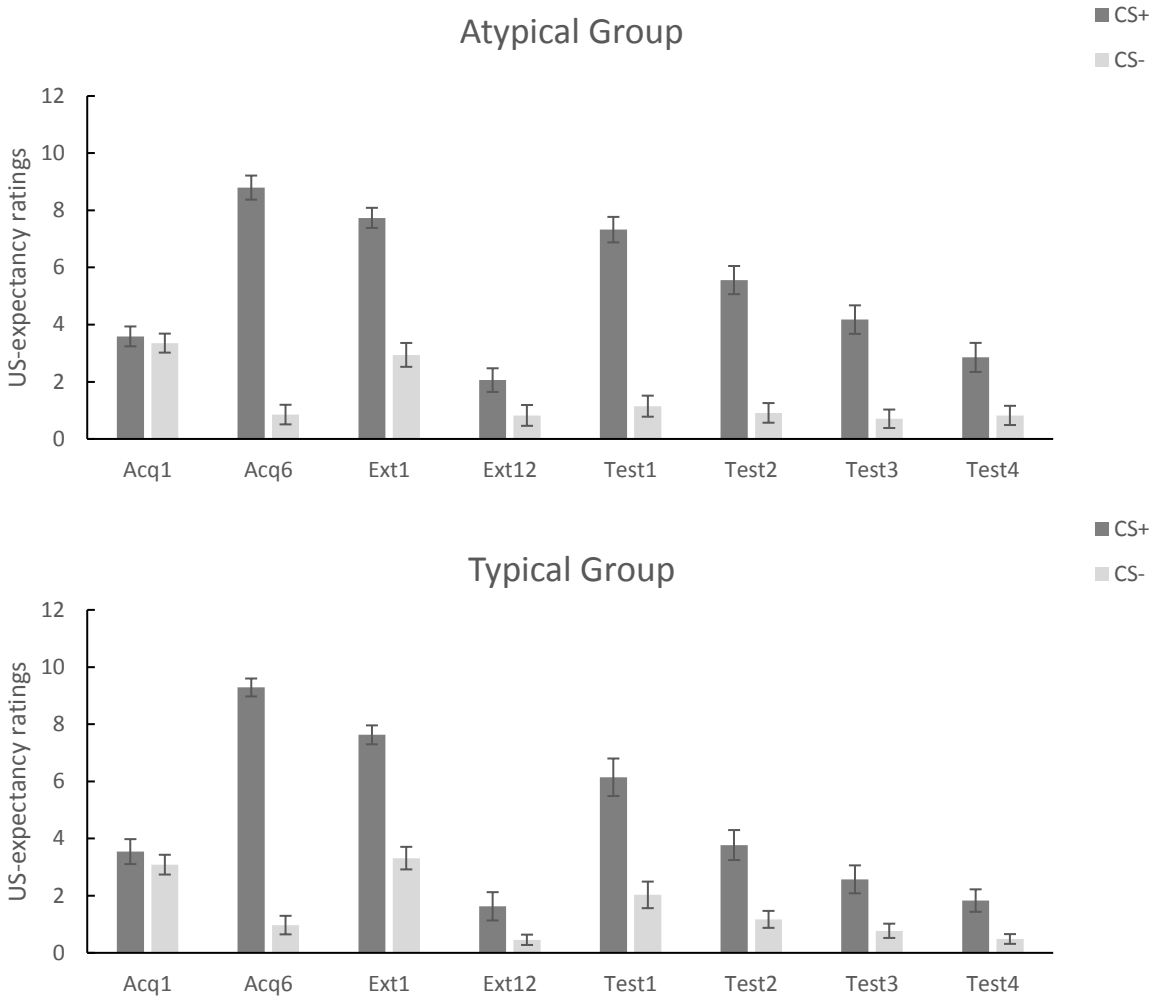

Figure 2. Mean US-expectancy ratings per Group and CS on the first acquisition trial (Acq1), the last acquisition trial (Acq6), the first extinction trial (Ext1), the last extinction trial (Ext12) and four test trials (Test1-4). Error bars represent standard errors.

Acquisition and generalization of acquisition. Acquisition learning was tested by comparing the first and the last trial in a 2 (Trial: acq1, acq6) × 2 (CS: CS+, CS-) × 2 (Group: typical, atypical) repeated measures analysis of variance (RM ANOVA). There was a significant main effect of CS, $F(1, 67) = 212.88$, $p < .001$, $\eta^2_p = .76$, and a significant CS × Trial interaction, $F(1, 67) = 160.72$, $p < .001$, $\eta^2_p = .71$. The CS × Trial interaction indicates that throughout acquisition training participants learned to discriminate between the CS+ and the CS-. The absence of a significant CS × Trial × Group interaction suggests that this learning effect was similar in both groups, $F(1, 67) = 0.02$, $p = .902$, $\eta^2_p = 0$. In addition, we performed

two separate 2 (CS: CS+, CS-) × 2 (Group: typical, atypical) RM ANOVAs on the first and last trial of acquisition. On the first acquisition trial, there was no significant difference in US-expectancies between the CS+ (M = 3.57, SD = 2.31) and CS- (M = 3.22, SD = 1.98), $F(1, 67) = 0.87$, $p = .354$, $\eta^2_p = .01$ (i.e., no main effect of CS). On the last acquisition trial, US-expectancies were significantly higher for the CS+ (M = 9.04, SD = 2.17) compared to the CS- (M = 0.91, SD = 1.95), $F(1, 67) = 301.29$, $p < .001$, $\eta^2_p = .82$.

The generalization of acquisition was tested by a 2 (CS: CS+_GS1, CS-_GS1) × 2 (Group: typical, atypical) RM ANOVA on the first extinction trial. The discrimination between the CS+_GS1 (M = 7.68, SD = 1.99) and CS-_GS1 (M = 3.13, SD = 2.38) was significant, evidenced by a main effect of CS, $F(1, 67) = 107.82$, $p < .001$, $\eta^2_p = .62$. The CS × Group interaction did not reach significance, $F(1, 67) = 0.48$, $p = .49$, $\eta^2_p = 0.01$. These results provide evidence for generalization of acquisition learning to the GS1 that was not different for the two groups.

In additional analyses, we tested the generalization of acquisition by comparing the last trial of acquisition with the first extinction trial in a 2 (Trial: acq6, ext1) × 2 (CS: CS+/CS+_GS1, CS-/CS-_GS1) × 2 (Group: typical, atypical) RM ANOVA. This resulted in a significant main effect of CS, $F(1, 67) = 361.77$, $p < .001$, $\eta^2_p = .84$, a significant CS × Trial interaction, $F(1, 67) = 33.72$, $p < .001$, $\eta^2_p = .34$, and a non-significant CS × Trial × Group interaction, $F(1, 67) = 0.48$, $p = .49$, $\eta^2_p = 0.01$. These results lead to the same conclusions as those reached in the analysis without controlling for the last acquisition trial. Although not problematic for the purpose of our study, it is of note that the significant CS × Trial interaction additionally suggests that the generalization of acquisition learning to the extinction GS was not complete: the CS+/CS- discrimination at the end of acquisition was larger than the CS+_GS1/CS-_GS1 discrimination at the beginning of extinction.

Extinction and generalization of extinction. Extinction learning was tested by comparing the first with the last extinction trial in a 2 (Trial: ext1, ext12) × 2 (CS: CS+_GS1, CS-_GS1) × 2 (Group: typical, atypical) RM ANOVA. There was a significant main effect of CS, $F(1, 67) = 85.42$, $p < .001$, $\eta^2_p = .56$, and a significant CS × Trial interaction, $F(1, 67) = 69.88$, $p < .001$, $\eta^2_p = .51$. This significant interaction suggests that the difference in responding to the CS+_GS1 and CS-_GS1 decreased across the extinction trials. However, US-expectancies for the CS+_GS1 (M = 1.84, SD = 2.68) were still significantly higher than for CS-_GS1 (M = 0.64, SD = 1.67) at the end of extinction, $F(1, 67) = 17.66$, $p < .001$, $\eta^2_p = .21$,

which suggests that the discrimination between the CS+_GS1 and CS-_GS1 did not disappear completely. The CS × Trial × Group interaction was not significant, $F(1, 67) = 0.27$, $p = .61$, $\eta^2_p = 0$, indicating no differences between the typical and atypical group in extinction learning.

To test for return of fear, a 2 (CS: CS+_GS2, CS-_GS2) × 2 (Group: typical, atypical) RM ANOVA was performed on the first test trial. The CS × Group interaction was significant, $F(1, 67) = 4.64$, $p = .035$, $\eta^2_p = .072$. This confirms that the typical group showed less discrimination between the CS+_GS2 ($M = 6.14$, $SD = 3.89$) and the CS-_GS2 ($M = 2.03$, $SD = 2.75$) than the atypical group (CS+_GS2: $M = 7.32$, $SD = 2.60$; CS-_GS2: $M = 1.14$, $SD = 2.15$). A 4 (Trial: test1, test 2, test 3, test 4) × 2 (CS; CS+_GS2, CS-_GS2) × 2 (Group: typical, atypical) RM ANOVA was used to examine the subsequent test trials. There was a significant CS × Group interaction, $F(1, 67) = 6.24$, $p = .015$, $\eta^2_p = .09$. This indicates that the effect of the instructions was persistent. The between-group difference between the CS+_GS2 and CS-_GS2 stands when taking all test trials into account: the typical group showed less discrimination between the CS+_GS2 (test2: $M = 3.77$, $SD = 3.11$; test3: $M = 2.57$, $SD = 2.89$; test4: $M = 1.83$, $SD = 2.33$) and CS-_GS2 (test2: $M = 1.17$, $SD = 1.76$; test3: $M = 0.77$, $SD = 1.48$; test4: $M = 0.49$, $SD = 1.01$) throughout test compared to the atypical group (CS+_GS2: test2: $M = 5.56$, $SD = 2.87$; test3: $M = 4.18$, $SD = 2.91$; test4: $M = 2.85$, $SD = 2.97$; CS-_GS2: test2: $M = 0.91$, $SD = 2.01$; test3: $M = 0.71$, $SD = 1.88$; test4: $M = 0.82$, $SD = 1.96$).

We performed additional analyses to control for the CS+_GS1/CS-_GS1 difference at the end of extinction training. The last extinction trial was compared with the first test trial using a 2 (Trial: ext12, test1) × 2 (CS: CS+_GS1/GS2, CS-_GS1/GS2) × 2 (Group: typical, atypical) RM ANOVA. This revealed a significant Trial × CS interaction, $F(1, 67) = 59.37$, $p < .001$, $\eta^2_p = .47$, and a Trial × CS × Group interaction that approached significance, $F(1, 67) = 3.81$, $p = .055$, $\eta^2_p = .05$. The Trial × CS × Group interaction points to a differential increase in US-expectancies from extinction to test that was influenced by the typicality instruction in the predicted direction (see Figure 2). However, it is of note that when taking into account the last acquisition trial, the effects of the typicality instruction only approached the $\alpha = .05$ level of significance. When the last extinction trial was compared with the mean of the four test trials in a 2 (Trial: ext12, mean test1-4) × 2 (CS: CS+_GS1/GS2, CS-_GS1/GS2) × 2 (Group: typical, atypical) RM ANOVA, the Trial × CS × Group interaction reached significance, $F(1, 67) = 5.34$, $p = .024$, $\eta^2_p = .07$.

SCR

Three participants were excluded due to technical error leaving 66 final participants ($N_{atypical}$ = 32; $N_{typical}$ = 34). Figure 3 indicates mean SCR amplitudes for the CSs and GSs, for both groups. Displayed are the first and last acquisition trials, the first and last extinction trials, and the four test trials.

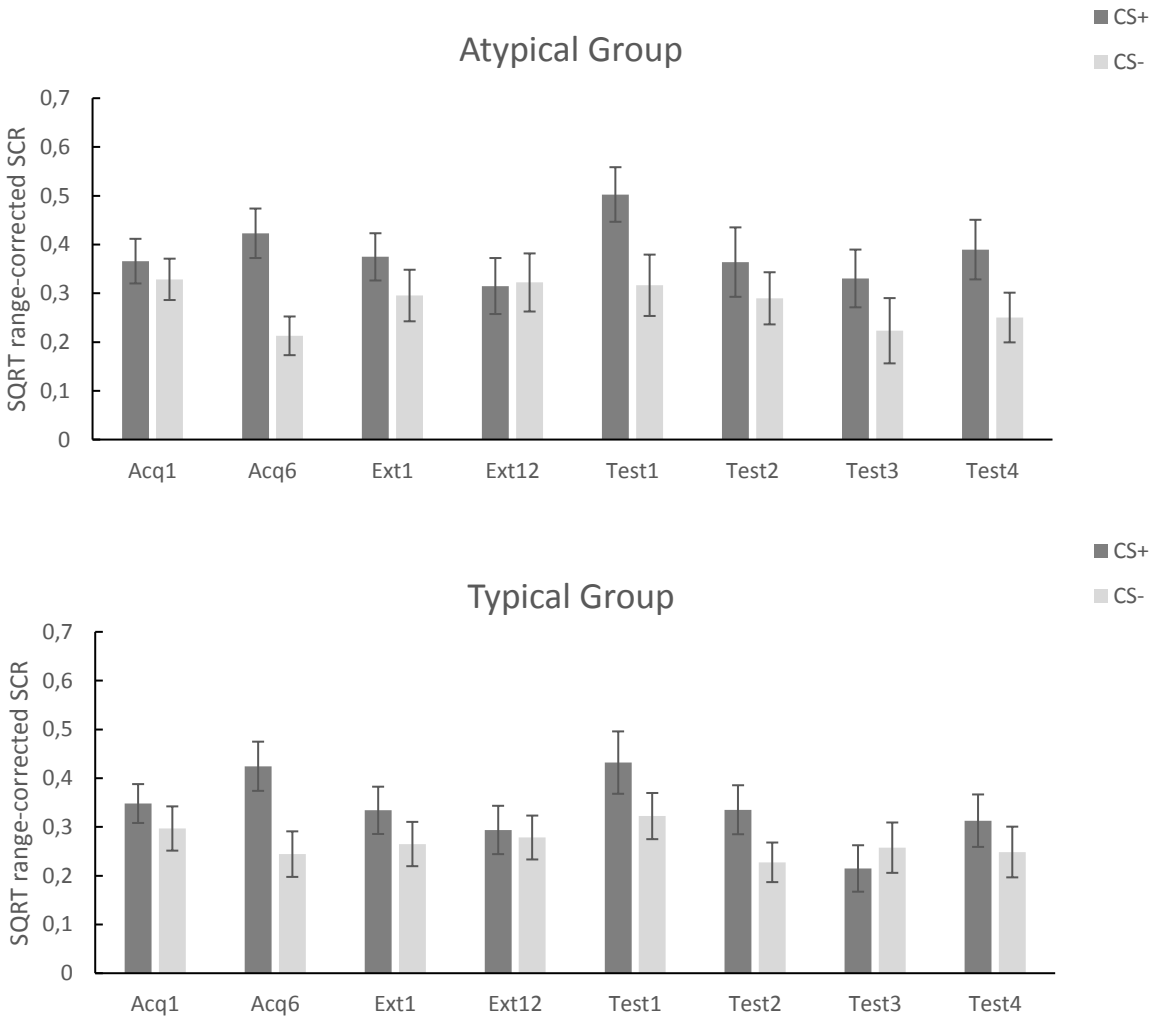

Figure 3. SQRT range-corrected mean SCR per Group and CS on the first acquisition trial (Acq1), the last acquisition trial (Acq6), the first extinction trial (Ext1), the last extinction trial (Ext12) and four test trials (Test1-4). Error bars represent standard errors.

Acquisition and generalization of acquisition. A 2 (Trial: acq1, acq6) × 2 (CS: CS+, CS-) × 2 (Group: typical, atypical) RM ANOVA on the first and last trial of acquisition showed

a significant main effect of CS, $F(1, 64) = 28.85$, $p < .001$, $\eta^2_p = .31$, and a significant CS × Trial interaction, $F(1, 64) = 10.54$, $p < .01$, $\eta^2_p = .14$. The CS × Trial interaction suggests that a CS+/CS- discrimination emerged throughout acquisition. This analysis did not yield a significant CS × Trial × Group interaction, $F(1, 64) = 0.22$, $p = .64$, $\eta^2_p = 0$. This indicates that there were no differences between the typical and atypical group in the course of acquisition training. In addition, two separate 2 (CS: CS+; CS-) × 2 (Group: typical, atypical) RM ANOVAs were performed on the first and last acquisition trials. At the end of acquisition, participants responded higher to the CS+ ($M = 0.42$, $SD = 0.29$) than to the CS- ($M = 0.23$, $SD = 0.25$). This was confirmed by a significant main effect of CS on the last acquisition trial, $F(1, 64) = 29.60$, $p < .001$, $\eta^2_p = .32$. This CS+/CS- discrimination was not present at the start of acquisition (CS+: $M = 0.36$, $SD = 0.24$; CS-: $M = 0.35$, $SD = 0.33$), as evidenced by a non-significant main effect of CS on the first acquisition trial, $F(1, 64) = 2.49$, $p = .12$, $\eta^2_p = .04$.

In a 2 (CS: CS+_GS1, CS-_GS1) × 2 (Group: typical, atypical) RM ANOVA on the first extinction trial, a marginally significant main effect of CS was found, $F(1, 64) = 3.79$, $p = .056$, $\eta^2_p = .06$. Although less pronounced than in the US-expectancy ratings, this result suggests that on the first extinction trial participants discriminated between the CS+_GS1 and CS-_GS1 (CS+_GS1: $M = 0.35$, $SD = 0.28$; CS-_GS1: $M = 0.28$, $SD = 0.28$). This indicates that there was generalization of acquisition learning to the extinction GS. The non-significant CS × Group interaction implies that the generalization of acquisition was similar in both groups, $F(1, 64)= 0.09$, $p = .76$, $\eta^2_p = 0$. In addition, comparing the last trial of acquisition with the first extinction trial in a 2 (Trial: acq6, ext1) × 2 (CS: CS+/CS+_GS1; CS-/CS-_GS1) × 2 (Group: typical, atypical) RM ANOVA showed a significant main effect of CS, $F(1, 64) = 24.97$, $p < .001$, $\eta^2_p = .28$, and a significant CS × Trial interaction, $F(1, 64) = 7.05$, $p = .01$, $\eta^2_p = .10$. No significant CS × Trial × Group interaction was found, $F(1, 64) = 0.01$, $p = .93$, $\eta^2_p = 0$. These results point to the same conclusions reached by the analysis without controlling for the last acquisition trial. Similar to the US-expectancy results, the significant CS × Trial interaction indicates that the generalization of acquired SCR to the extinction GS was not complete.

Extinction and generalization of extinction. A 2 (Trial: ext1, ext12) × 2 (CS: CS+_GS1, CS-_GS1) × 2 (Group: typical, atypical) RM ANOVA comparing the first and last extinction trial revealed no main effect of CS, $F(1, 64) = 1.58$, $p = .21$, $\eta^2_p = 0.02$ and no CS × Trial interaction, $F(1, 64) = 1.63$, $p = .21$, $\eta^2_p = .03$. The lack of CS × Trial interaction indicates that there was no adequate extinction learning. The CS × Trial × Group interaction was not

significant, $F(1, 64) = 0.12$, $p = .73$, $\eta^2_p = 0$. This suggests that there were no differences between the typical and atypical group in SCR during the extinction phase. At the same time, two separate 2 (CS: CS+_GS1, CS-_GS1) × 2 (Group: typical, atypical) RM ANOVAs on the first and last trial of extinction showed that the discrimination between the CS+_GS1 and CS-_GS1 was marginally significant at the start of extinction (cf. supra), but disappeared completely by the end of extinction (CS+_GS1: $M = 0.30$, $SD = 0.30$; CS-_GS1: $M = 0.30$, $SD = 0.30$), $F(1, 64) = 0$, $p = .99$, $\eta^2_p = 0$. Based on the above results, we have to conclude that there is no unequivocal evidence for extinction learning in SCR. The findings on return of fear reported below therefore require cautious interpretation.

A 2 (CS: CS+_GS2, CS-_GS2) × 2 (Group: typical, atypical) RM ANOVA on the first test trial was performed to test for return of fear. This resulted in a significant main effect of CS, $F(1, 64) = 11.14$, $p < .01$, $\eta^2_p = .15$. Hence, whereas discrimination between the CS+_GS1 and CS-_GS1 was not significant on the last extinction trial (cf. supra), it was for the CS+_GS2 and CS-_GS2 on the first test trial (CS+_GS2: $M = 0.47$, $SD = 0.35$; CS-_GS2: $M = 0.32$, $SD = 0.31$). Group differences in CS+_GS2/CS-_GS2 discrimination on the first test trial are in the predicted direction (see Figure 3), but the CS × Group interaction was not significant, $F(1, 64) = 0.75$, $p = .39$, $\eta^2_p = .01$ (typical group: CS+_GS2: $M = 0.43$, $SD = 0.37$; CS-_GS2: $M = 0.32$, $SD = 0.28$; atypical group: CS+_GS2: $M = 0.50$, $SD = 0.32$; CS-_GS2: $M = 0.32$, $SD = 0.36$). These results provide evidence for the return of fear in SCR when presented with the CS+_GS2. However, this return in SCR was not influenced by the instructions about typicality.

Similar conclusions are reached with a 2 (Trial: ext12, test1) × 2 (CS: CS+_GS1/GS2, CS-_GS1/GS2) × 2 (Group: typical, atypical) RM ANOVA comparing the last extinction trial with the first test trial. This resulted in a significant main effect of CS, $F(1, 64) = 6.29$, $p = .02$, $\eta^2_p = .09$, and a significant Trial × CS interaction, $F(1, 64) = 5.71$, $p = .02$, $\eta^2_p = .08$. However, the CS × Trial × Group interaction was not significant, $F(1, 64) = 0.55$, $p = .46$, $\eta^2_p = .01$. In addition, no effects of Group were found when the mean of the four test trials was included in a 2 (Trial; ext12, mean test1-4) × 2 (CS: CS+_GS1/GS2, CS-_GS1/GS2) × 2 (Group: typical, atypical) RM ANOVA that also included the last extinction trial, $F(1, 64) = 0.65$, $p = .43$, $\eta^2_p = .01$ (i.e., no CS × Trial × Group interaction).

Conclusions and Discussion

The present experiment investigated whether information about the typicality of the extinction stimulus can affect the generalization of reduced fear to a novel exemplar. After one stimulus

(CS+) was paired with an electric shock (US), participants were presented with another stimulus (CS+_GS1) without the US. This stimulus was a perceptually similar member of the same family as the CS+. Subsequently, we tested for the generalization of extinction learning to a second generalization stimulus (CS+_GS2) from the CS+ family. Between extinction and test, half of the participants were instructed that the extinction stimulus was a typical exemplar of the CS+ family (typical group), while the other half was instructed that the extinction stimulus was an atypical exemplar (atypical group). US-expectancy ratings and SCR were included as dependent variables.

Similar to previous findings demonstrating return of fear caused by stimulus change after extinction (e.g., Barry et al., 2016; Boddez et al., 2012, Vervliet et al., 2005), both groups showed an increase in conditioned US-expectancies and SCR when presented with the novel exemplar (CS+_GS2) after extinction training. However, in line with our hypothesis, participants in the typical group showed less return of conditioned US-expectancies when tested with the novel exemplar than participants in the atypical group.

The pattern in SCR was similar but did not reach significance. Importantly, we did not obtain unequivocal support for extinction in SCR either. Since extinction of fear can be considered a prerequisite for reliably testing its return, SCR results require cautious interpretation. It is of note that relatively less clear patterns in SCR are reported often and are regularly attributed to larger measurement error in SCR (e.g., Boddez, et al., 2013; Haesen & Vervliet, 2015; Schultz, Balderston, Geiger, & Helmstetter, 2013). Limited statistical power of our study might have resulted in difficulties to detect the subtle effects of the instruction in the SCR measure (e.g., Ahmed & Lovibond, 2015, but see Raes, De Houwer, Verschuere, & De Raedt, 2011). Although US-expectancy has sometimes been considered an inferior measure due to its subjective character compared to more ‘objective’ physiological indices of fear such as SCR, a systematic evaluation by Boddez et al. (2013) suggests that it is a robust and valid measure in human fear conditioning research. In particular, fear conditioning research relying on the US-expectancy measure has shown sufficient face validity, diagnostic validity, predictive validity and construct validity with respect to anxiety disorders (Boddez et al., 2013).

Several aspects of our study require further consideration. First, residual discrimination between the CS+_GS1 and CS-_GS1 was observed in the US-expectancy ratings on the last extinction trial. Such residual discrimination at the end of extinction has been argued to be consistent with the residual anxiety that is often observed at the end of exposure therapy (e.g., Vervliet et al., 2005). Because of this residual discrimination, we analyzed the US-expectancy

ratings in two ways; we included the last extinction trial in one model, but not in another. When not including the last extinction trial, the instructions significantly affected the return of fear, as was initially hypothesized. However, when including the last extinction trial in the analysis, the effect of the typicality instruction only approached significance. At the same time, results were again significant when comparing the last extinction trial with the mean of the four test trials.

Second, we did not include a control condition in which participants received no information about the typicality of the extinction stimulus. Our study therefore does not allow distinguishing between whether emphasizing the typical character of the extinction stimulus enhances extinction generalization or whether emphasizing its atypicality hampers extinction generalization. In case of the former, the recommendation for clinical practice would be to include interventions about how the exposure stimulus is a typical exemplar and resembles other stimuli of the feared category. In case of the latter, it might be recommended to refrain from any information or statements that relate to the unique or atypical character of the exposure stimulus. In the treatment of dog phobia, for instance, giving a name to the dog or providing any other specific (comforting) information about the dog (e.g., "it's the dog of a friend") might contribute to the patient's belief that the dog used during treatment is not representative for other dogs. A third and at least as plausible option is that both the former and the latter are true. In that case, a combination of both emphasizing the typical character and avoiding statements that are related to the atypical character of the exposure exemplar might promote the generalization to other exemplars. Notably, these recommendations for clinical practice can also be applied to other anxiety problems. For instance, in the case of exposure treatment for social phobia, therapists might provide information about the typical way of responding of the interaction partner and refrain from information that might suggest its atypicality (e.g., refrain from saying things like "this was a friendly person"). Similarly, for a patient with obsessive-compulsive disorder who is exposed to holding a knife in the presence of a loved one, it can be recommended that the therapist emphasizes that the used knife is a typical exemplar (e.g., equally potentially dangerous as other knives) and that the present loved one is representative for other loved ones. However, a legitimate question is whether clinically anxious individuals are equally sensitive to this kind of verbal information. On top of that, it could be the case that instructions have higher credibility in a controlled experimental context than in real life. The experimenter could indeed be considered a highly credible source for future behavior of Fribbles in the experimental context, whereas people might, for example, not as readily believe

what their therapist has to say about future behavior of dogs. An obvious first step for future research, therefore, would be to verify that the effectiveness of our typicality manipulation also holds for clinically anxious individuals and for real-life fears.

Our results are complementary to the study of Dunsmoor and Murphy (2014), in which the effect of stimulus typicality on the generalization of fear acquisition was investigated. However, whereas Dunsmoor and Murphy (2014) directly manipulated the typicality of the stimulus by presenting typical (e.g., rabbit) or atypical (e.g., armadillo) exemplars of existing categories (e.g., mammals), we manipulated stimulus typicality through verbal instructions. A target for future research is therefore to test the effects of manipulating the typicality of the extinction (exposure) stimulus itself on extinction generalization (e.g., using existing categories as in Dunsmoor & Murphy, 2014). If the results of such study parallel our results, this might have implications for clinical practice. In particular, it would suggest that a typical exemplar of the feared category (e.g., in the case of dog phobia: a Labrador) is preferred above an atypical exemplar (e.g., a Chinese crested dog) in clinical exposure therapy. The advantage of using verbal instructions, however, is that it is a straightforward and feasible strategy that can easily be implemented in clinical practice (cf. supra). Selecting a typical stimulus requires that the therapist has a wide range of stimuli available. In addition, individual patients might differ from each other and over time in their judgement about the typicality of a particular stimulus, due to for instance a different amount of pre-exposure to the stimulus (e.g., Barsalou, 1985; Novick, 2003).

In summary, the present study provides evidence for the influence of verbal information about the typicality of the extinction stimulus on extinction generalization to a new exemplar.

References

Ahmed, O., & Lovibond, P. F. (2015). The impact of previously learned feature-relevance on generalization of conditioned fear in humans. *Journal of Behavior Therapy and Experimental Psychiatry, 46,* 59-65. doi:10.1016/j.jbtep.2014.08.001

Barry, T. J., Griffith, J. W., De Rossi, S., & Hermans, D. (2014). Meet the Fribbles: novel stimuli for use within behavioural research. *Frontiers in Psychology, 5,* 1-8. doi:10.3389/fpsyg.2014.00103

Barry, T. J., Griffith, J. W., Vervliet, B., & Hermans, D. (2016). The role of stimulus specificity and attention in the generalization of extinction. *Journal of Experimental Psychopathology, 7,* 143-152. doi:10.5127/jep.048615

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11,* 629-654.

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,* 201-206. doi:10.1016/j.jbtep.2012.08.003

Boddez, Y., Bennett, M., P., van Esch, S., & Beckers, T. (2016). Bending rules: The shape of the perceptual generalization gradient is sensitive to inference rules. *Cognition and Emotion*.

Boddez, Y., Callaerts-Vegh, Z., Vervliet, B., Baeyens, F., D'Hooge, R., Hermans, D., & Beckers, T. (2012). Stimulus generalization and return of fear in C57BL/6J mice. *Frontiers in Behavioral Neuroscience, 6,* 1-5. doi:10.3389/fnbeh.2012.00041

Bouton, M. E. (2002). Context ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976-986. doi:10.1016/S0006-3223(02)01546-9

Craske, M. G., & Mystkowski, J. L. (2006). Exposure therapy and extinction: Clinical studies. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and learning: From basic processes to clinical implications* (pp. 217-233). Washington, DC: American Psychiatric Association.

Dawson, M. E., Schell, A. M., & Fillion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G., Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 200-224). Cambridge: Cambridge University Press.

de Clercq, A., Verschuere, B., de Vlieger, P., & Crombez, G. (2006). Psychophysiological

analysis (PSPHA): a modular script-based program for analyzing psychophysiological data. *Behavior Research Methods, 38*, 504-510. doi:10.3758/BF03192805

Dubin, W. J., & Levis, D. J. (1975). Generalization of extinction gradients: A systematic analysis. *Journal of Experimental Psychology, 100*, 403-412.

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus typicality determines how broadly fear is generalized. *Psychological Science, 25*, 1816-1821. doi:10.1177/0956797614535401

Fonteyne, R., Vervliet, B., Hermans, D., Baeyens, F., & Vansteenwegen, D. (2010). Exposure to the context and removing the unpredictability of the US: Two methods to reduce contextual anxiety compared. *Biological Psychology, 85*, 361-369. doi:10.1016/j.biopsycho.2010

Haesen, K., & Vervliet, B. (2015). Beyond extinction: Habituation eliminates conditioned skin conductance across contexts. *International Journal of Psychophysiology, 98,* 529-534. doi:10.1016/j.ijpsycho.2014.11.010

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

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: retrospective revaluation induced by experience and instruction. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*, 97-106. doi:10.1037/0278-7393.29.1.97

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: a proposal for standardization. *Psychophysiology, 8*, 656-672.

McNally, R. J. (2016). The legacy of Seligman's "Phobias and preparedness" (1971). *Behavior Therapy, 47*, 585-594. doi:10.1013/j.beth.2015.08.005

Mertens, G., & De Houwer, J. (2016). The impact of a context switch and context instructions on the return of verbally conditioned fear. *Journal of Behavior Therapy and Experimental Psychiatry, 51,* 10-18. doi:10.1016/j.jbrep.2015.11.001

Muris, P., & Field, A. (2010). The role of verbal threat information in the development of childhood fear. "Beware the Jabberwock!". *Clinical Child and Family Psychology review, 13*, 129-150. doi:10.1007/s10567-010-0064-1

Norton, P. J., & Price, E. C. (2007). A meta-analytic review of adult cognitive-behavioral treatment outcome across the anxiety disorder. *The Journal of Nervous and Mental Disease, 195*, 521-531. doi:10.1097/01.nmd.0000253843.70149.9a

Novick, L. R. (2003). At the forefront of thought: The effect of media exposure on airplane typicality. *Psychonomic Bulletin & Review, 10*, 971-974. doi:10.3758/BF03196560

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after Pavlovian,

observational and instructed fear. *Psychological Science, 15,* 822-828. doi:10.1111/j.0956-7976.2004.00762.x

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185-200.

Pineles, S. L., Orr, M. R., & Orr, S. P. (2009). An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. *Psychophysiology, 46,* 984-995. doi:10.1111/j.1469-8986.2009.00852.x.

Raes, A. K., De Houwer, J., Verschuere, B., & De Raedt, R. (2011). Return of fear after retrospective inferences about the absence of an unconditioned stimulus during extinction. *Behaviour Research and Therapy, 49,* 212-218. doi:10.1016/j.brat.2010.12.004

Rowe, M. K., & Craske, M. G. (1998). Effects of varied-stimulus exposure training on fear reduction and return of fear. *Behaviour Research and Therapy, 36,* 719-734. doi:10.1016/S00057967(97)10017-1

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

Schultz, D. H., Balderston, N. L., Geiger, J. A., & Helmstetter, F. J. (2013). Dissociation between implicit and explicit responses in postconditioning UCS revaluation after fear conditioning in humans. *Behavioral Neuroscience, 127,* 357-368. doi:10.1037/a0032742

Vervoort, E., Vervliet, B., Bennett, M., & Baeyens, F. (2014). Generalization of human fear acquisition and extinction within a novel arbitrary stimulus category. *PLoS One, 9,* e96569. doi:10.1371/journal.pone.0096569

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: State of the art. *Annual Review of Clinical Psychology, 9,* 215-248. doi:10.1146/annurev-clinpsy-050212-185542

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy, 43,* 357-371. doi:10.1016/j.brat.2004.02.005

Vervliet, B., Vansteenwegen, D., & Eelen, P. (2006). Generalization gradients for acquisition

and extinction in human contingency learning. *Experimental Psychology, 53*, 132-142. doi:10.1027/1618-3169.53.2.132

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

Chapter 3

Modeling hierarchical versus random exposure schedules in Pavlovian fear extinction: No evidence for differential fear outcomes

Based on:

Scheveneels, S., Boddez, Y., Vervliet, B., & Hermans, D. (under review). Modeling hierarchical versus random exposure schedules in Pavlovian fear extinction: No evidence for differential fear outcomes. *Behavior Therapy*.

Abstract

In exposure therapy, the client can either be confronted with the fear-eliciting situations in a hierarchical way (i.e., moving gradually from less to more fear-eliciting situations) or in a random way (i.e., being exposed to less and more fear-eliciting situations randomly). In the current study we developed a procedure to investigate the effects of hierarchical versus random exposure on long-term fear responding in the laboratory. Using a fear conditioning procedure, one stimulus (CS+) was paired with an electric shock (US), whereas another stimulus was not paired with the shock (CS-). The next day, participants underwent extinction training including presentations of the CS-, CS+ and a series of morphed stimuli between the CS+ and CS-. In the hierarchical extinction condition (HE; $N = 32$), participants were first presented with the CS-, subsequently with the morph most similar to the CS-, then with the morph most similar to that one, and so forth, until reaching the CS+. In the random extinction condition (RE; $N = 32$), the same stimuli were presented but in a random order. Fear responding to the CS+, CS- and a new generalization stimulus (GS) was measured on the third day. Higher expectancy violation ($t(62) = -2.67$, $p = .01$), physiological arousal ($t(62) = -2.08$, $p = .04$) and variability in US-expectancy ratings ($t(62) = -2.25$, $p = .03$) were observed in the RE condition compared to the HE condition, suggesting the validity of this novel procedure. However, no differences between the RE and HE condition were found in fear responding as tested one day later, $F(1, 62) < 1$.

Modeling hierarchical versus random exposure schedules in Pavlovian fear extinction: No evidence for differential fear outcomes

Exposure is a key component in the treatment of anxiety and related disorders (e.g., Choy, Fyer, & Lipsitz, 2007; Sánchez-Meca, Rosa-Alcázar, Marín-Martínez, & Gómez-Conesa, 2010). In exposure therapy, the patient is confronted with fear-eliciting stimuli or situations. In guiding therapists how to conduct exposure therapy, traditional models of exposure such as Emotional Processing Theory (EPT) encourage the use of exposure hierarchies, whereby this confrontation takes place in a graduated manner (Foa, Huppert, & Cahill, 2006; Foa & Kozak, 1986). Typically, it is recommended to start the exposure with stimuli that elicit only little fear and to gradually progress towards the more distressing items. For example, a patient with fear of heights can first be exposed to standing on a 1st floor balcony with safety rails, moving stepwise towards the edge. Once fear has declined in that situation, one could repeat this on progressively higher balconies and on balconies without safety rails.

In contrast, recent suggestions based on Inhibitory Learning Theory (ILT) hold that the use of exposure hierarchies is not necessary and might even be detrimental for long-term fear reduction (Craske et al., 2008; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; Weisman & Rodebaugh, 2018). Instead, therapies based on ILT prescribe to conduct exposure in a variable way, going through the items of the hierarchy in a random order rather than in a linear fashion (Craske et al., 2014; Knowles & Olatunji, 2018). For example, having exposure on a different floor every time instead of gradually moving from lower to higher floors. In particular, random exposure might result in higher levels of variability in subjective fear and physiological arousal during treatment (e.g., Kircanski et al., 2012). Other research has demonstrated that such variability is associated with less fear responding at retest; perhaps because it facilitates memory consolidation (Brown, LeBeau, Chat, & Craske, 2017; Cain, Blouin, & Barad, 2004; Rescorla, 2006; Culver, Stoyanova, & Craske, 2012; Kircanski et al., 2012).

In addition, it has been argued that random exposure might allow for more expectancy violation[1], which according to ILT drives fear reduction during exposure therapy (Craske et al., 2008). Expectancy violation refers to the mismatch between the expected outcome (e.g., falling off the balcony and dying) and the actual outcome (e.g., not falling of the balcony and dying;

[1] It is important to note that the correctness of the theorem that the net amount of prediction error will be higher throughout the course of random (versus hierarchical) exposure treatment depends on certain parameters (e.g., concerning the amount of generalization between the presented feared stimuli). In other words, theoretically speaking, this will not always be the case.

Rescorla & Wagner, 1972). It is assumed that the stronger the mismatch, the more (long-term) fear reduction one obtains.

To date, two clinical studies have directly compared random exposure with a hierarchical approach. In a study by Lang and Craske (2000), participants fearful of heights had one hour of exposure divided into twelve 5-min exercises. In each exercise, participants were asked to approach the rail at a particular height. In the hierarchical group, participants practiced from the lowest to the highest floor in a hierarchical order and always approached the rail in the same way. Participants in the random group practiced at the different heights in a random order and varied in how they approached the rail. Moreover, the random group practiced at two different locations, whereas participants in the hierarchical group always practiced at the same location. Results revealed no differences in (long-term) treatment outcomes between both groups in the fear of heights. In a second study on this topic by Kircanski et al. (2012), participants with contamination fears were assigned to a random group or a hierarchical group. In the hierarchical group, participants were exposed to one item per session, progressing from the lowest to the highest item of the participant's fear hierarchy across the three sessions. So, they practiced with the least fear-provoking item during the first session, with the middle fear-provoking item during the second session and with the highest fear-provoking item during the third session. Within each session, the item was approached in the same gradual way and all exercises had the same duration. Participants in the random group were exposed to all items in a random order of difficulty during each session and for varying durations. Similar to the results of Lang and Craske (2000), no differences between both groups were found in the short- and long-term outcomes of exposure.

Hence, both clinical studies relevant to this topic did not find a difference in treatment outcome between random and hierarchical exposure. However, not only the order in which the feared stimuli were presented was manipulated in these studies, but also the number of stimuli and contexts as well as the timing of the exposure trials. In the current study, we aimed to compare the effects of hierarchical versus random exposure on fear responding using a fear conditioning paradigm, controlling for all other variables than the order in which the feared stimuli were presented. On the first day, participants went through an acquisition phase in which one stimulus was always paired with an electric shock (i.e., danger cue; CS+) and another stimulus was never paired with shock (i.e., safe cue; CS-). In the extinction phase, which took place on the second day, participants were presented with the CS+ and CS- and a set of morphed stimuli in between the CS+ and CS-. None of these cues was followed by shock and, crucially,

we manipulated the sequence in which the stimuli were presented. Participants in the hierarchical extinction (HE) condition were first presented with the CS- and then gradually proceeded to the CS+. So, after presenting the CS-, the morph that was most similar to it was presented. Subsequently, the most similar to that one was presented, etc. In the random extinction (RE) condition the same stimuli were presented but in a random order. On the third day, we tested fear responding to the CS+, the CS- and a new third stimulus which served as a generalization stimulus (GS). In line with the assumptions of ILT, we expected lower fear responding at test in the RE compared to the HE condition. Additionally, higher physiological arousal, higher levels of variability in US-expectancy ratings and higher expectancy violation were expected during extinction in the RE condition compared to the HE condition. It was predicted that, irrespective of condition, higher physiological arousal, higher variability in US-expectancy ratings and higher expectancy violation throughout extinction would be associated with lower fear responding at test.

Method

Participants

Sixty-four participants (M_{age} = 22.31; *SD* = 6.04; 53 females) took part in the experiment in exchange for course credit or payment. Half of the participants were assigned to the hierarchical extinction (HE) condition, the other half to the random extinction (RE) condition. Exclusion criteria were pregnancy, cardio-pulmonary conditions, psychiatric or neurological disorders (e.g., epilepsy) and wrist pain. All participants gave informed consent and the study was approved by the ethical committee of the Faculty of Psychology and Educational Sciences of the University of Leuven.

Apparatus and Stimuli

Conditioned stimuli. Two pictures of neutral male faces were used as conditioned stimuli (CS). It was counterbalanced which of the two faces functioned as the CS+ and which one as the CS-. The extinction stimuli (ES) were eight stepwise morphs between the two CSs (Figure 1; e.g., Leer et al., 2017; Lenaert et al., 2012). A third picture of another neutral male face served as the generalization test stimulus (GS) (Figure 1). All pictures were approximately 88

mm wide and 132 mm high (332 × 500 pixels) and were presented against a black background on a 19 inch Dell monitor (type P1911, resolution: 1440 × 900 at 60 Hz).

CS+	**ES1**	**ES2**	**ES3**	**ES4**	**ES5**	**ES6**	**ES7**	**ES8**	**CS-**	**GS**
CS-	**ES8**	**ES7**	**ES6**	**ES5**	**ES4**	**ES3**	**ES2**	**ES1**	**CS+**	**GS**

Figure 1. CS+, CS-, eight morphs that served as extinction stimuli (ES) and the generalization test stimulus (GS). CS+, CS- and ESs were counterbalanced.

Unconditioned stimulus. As an unconditioned stimulus (US), a 2 ms electrocutaneous stimulus was administered to the participant's right wrist. The US was delivered by a Digitimer DS7A constant current stimulator (Hertfordshire, UK), through a pair of V91-01 8-mm reusable Bilaney Ag/AgCl electrodes filled with K-Y jelly.

Measures

US-expectancy ratings. US-expectancies were rated by the participant on a trial-by-trial basis on an 11-point scale. The scale ranged from 0 = "certainly no shock" to 10 = "certainly shock". Participants were instructed to click a position on the scale that matched their expectancy using the computer mouse. The rating scale appeared onscreen 200 ms after stimulus onset. Participants had 7 s to respond. After 7 s or after participants gave their rating the scale disappeared.

Skin conductance response (SCR). Electrodermal responding was measured through a pair of disposable Biopac EL 507 electrodes (contact area = 95 mm^2) filled with isotonic paste and attached to the hypothenar site of the left hand palm. A Coulbourn LabLinc V Isolated Skin Conductance coupler (model V71-23, Coulbourn Instruments, Allentown, PA) applied a constant voltage of 0.5 Volts through these electrodes. The analog signal was digitized at 10

Hz by a NI PCI 3221 data acquisition card (National Instruments Corporation, Austin, Texas) from 2 s prior stimulus onset until 6 s after stimulus offset.

Procedure

The experiment consisted of three assessments on three subsequent days (i.e., Day 1, Day 2 and Day 3).

Day 1 started with general instructions and completion of the informed consent. Subsequently, the electrodes were attached and the intensity of the US was set to a level that was described as "definitely uncomfortable and demanding some effort to tolerate, but not painful" for each individual participant using a standard work-up procedure. The mean US intensity in this study was 21.91 mA (*SD* = 12.94). Following this, instructions about the experimental task were presented on the computer screen and orally repeated by the experimenter. Specifically, participants were instructed that pictures of faces would appear on the screen and that some of these faces could be followed by an electric shock. It was stated that participants' task was to predict the occurrence of the electric shock. Participants were then trained to use the rating scale in three practice trials. After each practice trial participants received feedback on whether or not the scale was filled in correctly. No stimuli were presented during these practice trials.

A schematic overview of the experimental phases is displayed in Table 1. On Day 1, the experimental task started with a pre-exposure phase in which the CS+ and CS- were presented twice without the US. This phase was added to reduce orienting responses in the skin conductance measure. The subsequent acquisition phase consisted of eight presentations of the CS+ followed by the US and eight CS- trials in absence of the US. Trial order was semi-randomized with no more than two consecutive trials with the same stimulus. To control for order-effects in both the pre-exposure phase and the acquisition phase, the CS+ was presented first for half of the participants in each condition and the CS- was presented first for the other half (e.g., Lovibond, 2003; Vervliet et al., 2005).

On Day 2 participants returned to the laboratory to continue with the extinction training. After attachment of the SCR and shock electrodes, participants were instructed that their task was the same as on Day 1 and to keep in mind what they had previously learned about the faces. In both groups the same set of stimuli was presented, but the sequence in which this happened was manipulated. In the HE condition, participants were first presented with the CS- and then

gradually moved from the CS- to the CS+ along the morphs. Each stimulus displayed in Figure 1 was presented twice, resulting in a total of 20 extinction trials (i.e., CS- → CS- → ES8 → ES8 → ES7 → ES7 → ES6 → ES6 → …). Participants in the RE condition were presented with the same stimuli but in a random order (e.g., ES6 → ES2 → ES7 → ES6 → CS- → ES3 → CS+ → ES8 → …). In this condition the stimulus sequence was determined randomly per individual participant. No USs were administered during extinction training.

On Day 3, the SCR and shock electrodes were again attached. Participants were told that their task was the same as previously and that they could use the information that they had previously learned. Subsequently, the CS+ and CS- were presented (in counterbalanced order), followed by the GS. This test cycle was repeated four times.

Table 1

Overview of the experimental phases

	Day 1		Day 2	Day3
	Pre-exposure	Acquisition	Extinction	Test
HE condition	CS+ (2) CS- (2)	CS+ (8) CS- (8)	CS- (2) → ES8 (2) → ES7 (2) →ES6 (2) → ES5 (2) → ES4 (2) → ES3 (2) → ES2 (2) → ES1 (2) → CS+ (2)	CS+ (4) CS- (4) GS (4)
RE condition	CS+ (2) CS- (2)	CS+ (8) CS- (8)	Random per participant: CS- (2); ES8 (2); ES7 (2); ES6 (2); ES5 (2); ES4 (2); ES3 (2); ES2 (2); ES1 (2); CS+ (2)	CS+ (4) CS- (4) GS (4)

Note. HE refers to hierarchical extinction, RE refers to random extinction. CSs, ESs (counterbalanced) and the GS are pictures of human faces. During acquisition '+' refers to the administration of the US and '-' to the absence of the US. During the other phases no USs are administered. The number of trials is indicated between parentheses.

In all experimental phases, each trial started with a 2 s blank screen to measure baseline skin conductance. Stimuli were presented for 8 s. If applicable, USs were delivered 7.5 s after stimulus onset. Inter-trial intervals (ITI) were set at an average of 8 s (range 7-9 s). Trial order,

stimulus presentation, ITI and registration of the dependent variables were controlled by Affect 4.0 (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010).

Results

US-expectancy ratings

Figure 2 displays the mean US-expectancy ratings for the CSs, ESs and GS in the HE and RE condition. The first and last acquisition trials, 20 extinction trials, and four test trials are presented. A Greenhouse-Geisser correction was applied when the assumption of sphericity was violated.

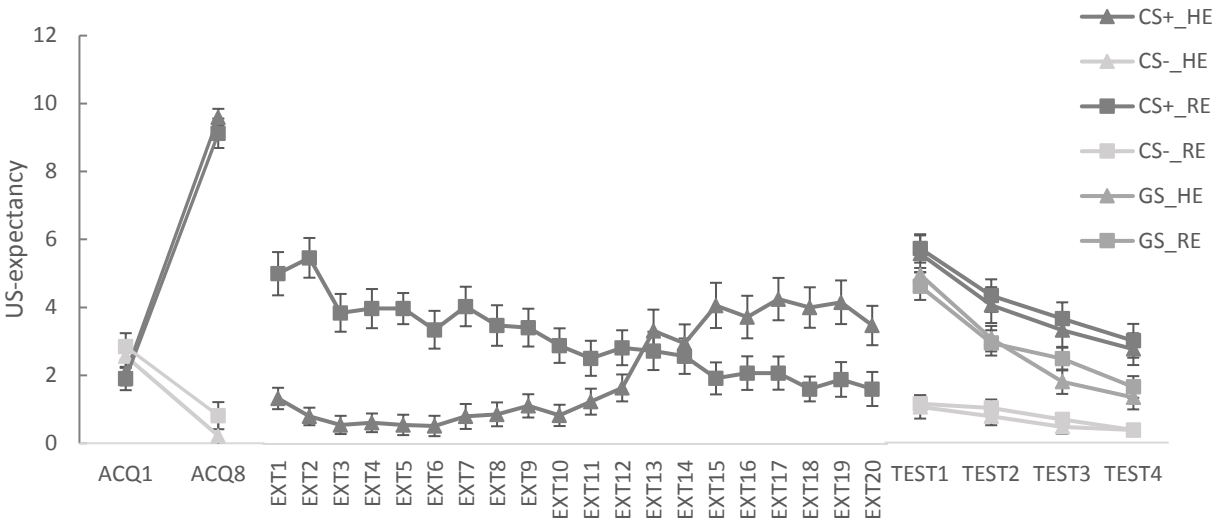

Figure 2. Mean US-expectancy ratings for the first and last acquisition trial, all extinction trials and all test trials per condition and CSs/GS. Error bars represent standard error of the means.

Acquisition phase. The left panel of Figure 2 suggests that participants learned to discriminate between the CS+ and CS- from the first to the last acquisition trial, with no differences between conditions. This was confirmed by a repeated measures Analysis of Variance (rmANOVA) with Trial (2 levels: acq1, acq8) and Stimulus (2 levels: CS+, CS-) as within-subjects factors and Condition (2 levels: HE, RE) as between-subjects factor. A main effect of Trial, $F(1, 62) = 121.41$, $p < .001$, $\eta^2_p = 0.66$, a main effect of Stimulus, $F(1, 62) = 524.95$, $p < .001$, $\eta^2_p = 0.89$, and a significant Trial × Stimulus interaction, $F(1, 62) = 821.83$, $p < .001$, $\eta^2_p = 0.93$, were found. The Trial × Stimulus interaction confirms the development of

CS+/CS- discrimination throughout the acquisition phase. As expected in this phase of the experiment, no effects of Condition were found, indicating that acquisition learning did not differ between the HE and RE condition.

Extinction phase. The middle panel of Figure 2 indicates a different course of the extinction phase in the RE and HE condition. A rmANOVA with Trial (20 levels: ext1 to ext20) as within-subjects factor and Condition (2 levels: HE, RE) as between-subjects factor revealed a significant Trial × Condition interaction, $F(19, 1178) = 16.83$, $p < .001$, $\eta^2_p = 0.21$, confirming a different course of the extinction phase in both conditions.

Test phase. The right panel of Figure 2 shows no differences in CS+/CS- discrimination between the RE and HE condition on the first test trial. This was confirmed by a rmANOVA with Stimulus (2 levels: CS+, CS-) as a within-subjects factor and Condition (2 levels: HE, RE) as a between-subjects factor. A significant main effect of Stimulus, $F(1, 62) = 106.08$, $p < .001$, $\eta^2_p = 0.63$, indicates that there was discrimination between the CS+ and CS- on the first test trial. However, no Stimulus × Condition interaction was found, $F(1, 62) = 0.01$, $p = .94$, $\eta^2_p = 0$. This result suggests that the CS+/CS- discrimination on the first test trial did not differ between the RE and HE condition. Similar results were found with regard to the GS/CS- discrimination on the first test trial. A rmANOVA on the first test trial with Stimulus (2 levels: GS, CS-) as a within-subjects factor and Condition (2 levels: HE, RE) as a between-subjects factor revealed a significant main effect of Stimulus, $F(1, 62) = 5.55$, $p < .05$, $\eta^2_p = 0.08$, but no Stimulus × Condition interaction, $F(1, 62) = 0.45$, $p = .50$, $\eta^2_p = 0.01$.

Additionally, we tested whether the course of the test phase differed between the HE and RE condition by conducting a rmANOVA with Stimulus (2 levels: CS+, CS-) and Trial (4 levels: test1, test2, test3, test4) as within-subjects factors and Condition (2 levels: HE, RE) as a between-subjects factor. A main effect of Trial, $F(3, 186) = 59.21$, $p < .001$, $\eta^2_p = 0.49$, and a main effect of Stimulus, $F(1, 62) = 89.39$, $p < .001$, $\eta^2_p = 0.59$, were found. However, the Trial × Stimulus × Condition interaction was not significant, $F(3, 186) = 0.06$, $p = .98$, $\eta^2_p = 0.01$, suggesting that the course of the test phase did not differ between both conditions.

Skin-conductance response (SCR)

Matlab was used to parse the data offline in a trial-by-trial fashion and locate means and maxima relative to CS occurrence. For each trial the mean values measured at baseline (i.e., 2 s prior to

stimulus onset) were subtracted from the maximum values measured in the 7.5 s after stimulus onset (Pineles, Orr, & Orr, 2009). Negative difference scores were recoded to zero. The resulting SCR amplitudes (µSiemens) were range-corrected with each participant's largest unconditioned response measured during US present trials (Lykken & Venables, 1971). To normalize the data, a square root transformation was subsequently applied (Dawson, Schell, & Fillion, 2000).

Figure 3 displays the range-corrected, square root transformed SCR for the CSs, ESs and GS in the HE and RE condition. The first and last acquisition trials, 20 extinction trials, and four test trials are displayed. A Greenhouse-Geisser correction was applied when the assumption of sphericity was violated.

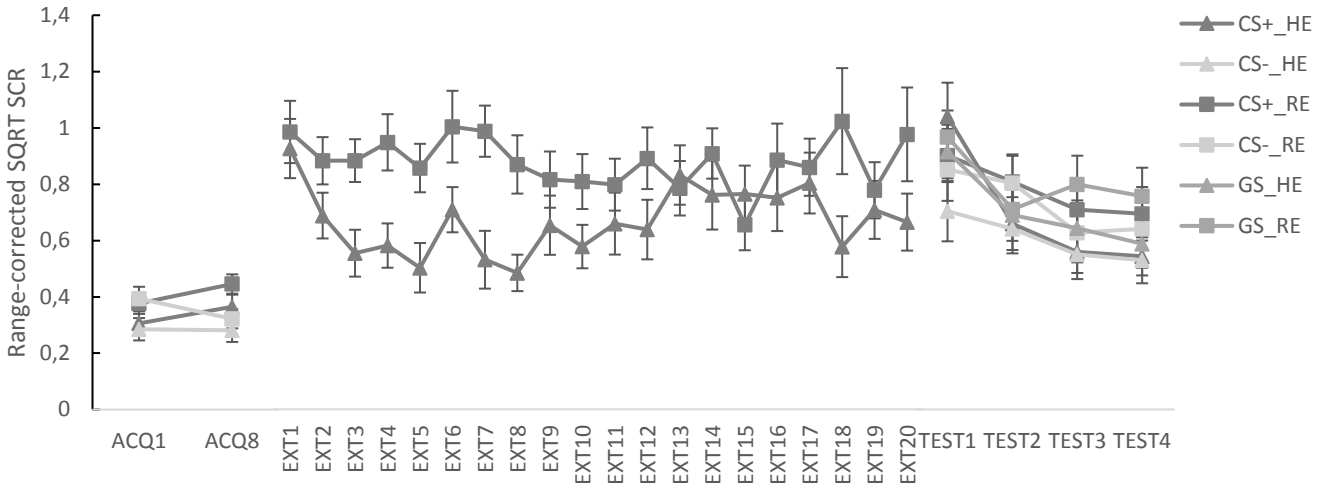

Figure 3. Range-corrected SQRT SCR for the first and last acquisition trial, all extinction trials and all test trials per condition and CS/GS. Error bars represent standard error of the means.

Acquisition phase. A rmANOVA with Trial (2 levels: acq1, acq8) and Stimulus (2 levels: CS+, CS-) as within-subjects factors and Condition (2 levels: HE, RE) as between-subjects factor revealed no main effect of Trial, $F(1, 62) = 0.25$, $p = .62$, $\eta^2_p = 0.01$, a significant main effect of Stimulus, $F(1, 62) = 4.25$, $p < .05$, $\eta^2_p = 0.06$, and a significant Trial × Stimulus interaction, $F(1, 62) = 4.08$, $p < .05$, $\eta^2_p = 0.06$. The significant Trial × Stimulus interaction suggests that participants started to discriminate between the CS+ and CS- over the course of acquisition. No effects of Condition were found in the acquisition phase.

Extinction phase. To test for group differences in the course of extinction, a rmANOVA with Trial (20 levels: ext1 to ext20) as within-subjects factor and Condition (2 levels: HE, RE) as between-subjects factor was performed. This resulted in a significant Trial × Condition interaction, $F(19, 1178) = 2.08$, $p < .05$, $\eta^2_p = 0.03$, suggesting that the RE and HE condition differed in the course of extinction.

Test phase. Group differences in fear responding on the first trial of the test phase were tested by a rmANOVA with Stimulus (2 levels: CS+, CS-) as a within-subjects factor and Condition (2 levels: HE, RE) as a between-subjects factor. This analysis revealed a main effect of Stimulus, $F(1, 62) = 4.74$, $p < .05$, $\eta^2_p = 0.07$, but no significant Stimulus × Condition interaction, $F(1, 62) = 2.59$, $p = .11$, $\eta^2_p = 0.04$. These results indicate no differences between the RE and HE condition with regard to CS+/CS- discrimination on the first test trial. In addition, with regard to the GS, a rmANOVA with Stimulus (2 levels: GS, CS-) as a within-subjects factor and Condition (2 levels: HE, RE) as a between-subjects factor revealed no significant main effect of Stimulus, $F(1, 62) = 0.17$, $p = .68$, $\eta^2_p = 0.01$, and no significant Stimulus × Condition interaction, $F(1, 62) = 1.81$, $p = .18$, $\eta^2_p = 0.03$. These results suggest that there is no GS/CS- discrimination on the first test trial in SCR, with no differences between the conditions.

Differences between the RE and HE condition in the course of the test phase were examined by conducting a rmANOVA with Stimulus (2 levels: CS+, CS-) and Trial (4 levels: test1, test2, test3, test4) as within-subjects factors and Condition (2 levels: HE, RE) as a between-subjects factor. This revealed a main effect of Trial, $F(3, 186) = 15.01$, $p < .001$, $\eta^2_p = 0.20$, no main effect of Stimulus, $F(1, 62) = 2.74$, $p = .10$, $\eta^2_p = 0.04$, and no significant Trial × Stimulus × Condition interaction, $F(3, 186) = 1.43$, $p = .24$, $\eta^2_p = 0.02$. These results indicate that the RE and HE group did not differ in the course of the test phase. With regard to GS/CS- discrimination, a rmANOVA with Stimulus (2 levels: GS, CS-) and Trial (4 levels: test1, test2, test3, test4) as within-subjects factors and Condition (2 levels: HE, RE) as a between-subjects factor was conducted. Results indicate a main effect of Trial, $F(3, 186) = 16.70$, $p < .001$, $\eta^2_p = 0.21$, no main effect of Stimulus, $F(1, 62) = 0.36$, $p = .55$, $\eta^2_p = 0.01$, and no significant Trial × Stimulus × Condition interaction, $F(3, 186) = 1.41$, $p = .24$, $\eta^2_p = 0.02$, suggesting that the test phase had a similar course in both conditions.

Expectancy violation, physiological arousal, variability throughout extinction and their correlation with fear responding at test

It was predicted that random extinction would result in higher *expectancy violation* throughout the extinction phase compared to hierarchical extinction. This was tested by comparing the sum of the US-expectancy ratings across all extinction trials between the RE and HE condition in an independent-samples t-test. Results confirm that the sum of the US-expectancy ratings was significantly higher in the RE condition (M = 61.69; SD = 5.16) compared to the HE condition (M = 41.16; SD = 5.70), $t(62) = -2.67$, $p = .01$.

In line with ILT, we predicted that individuals with higher expectancy violation during extinction would show less fear responding (i.e., CS+/CS- discrimination) on the first test trial. A significant positive correlation ($r = .27$, $p = .03$) was found between CS+/CS- discrimination on the first test trial and the sum of US-expectancies during extinction, suggesting that the higher the expectancy violation, the more discriminative fear responding in test. No significant correlation was found between the sum of the US-expectancies during extinction and differential SCR on the first test trial.

Furthermore, higher *physiological arousal*, as measured by the sum of the range-corrected, square root transformed SCR, was expected throughout extinction in the RE condition compared to the HE condition. An independent-samples t-test, $t(62) = -2.08$, $p = .04$, confirmed that the RE condition (M = 18.48; SD = 1.57) showed higher SCR throughout extinction compared to the HE condition (M = 14.04; SD = 1.46). Physiological arousal during extinction was, however, not associated with fear responding on the first test trial. No significant correlations were found between the sum of SCR across all extinction trials and discriminative fear responding on the first test trial in both US-expectancies and SCR.

Finally, we tested whether participants in the RE condition reported higher *variability* than those in the HE condition by calculating the standard deviation (SD) of the US-expectancy ratings across all extinction trials for each participant (e.g., Culver et al., 2012). An independent-samples t-test confirms that participants in the RE condition showed higher variability in their US-expectancy ratings (M = 2.81; SD = 0.12) than participants in the HE condition (M = 2.22; SD = 0.23), $t(62) = -2.25$, $p = .03$. In addition, a significant positive correlation ($r = .65$, $p < .001$) was found between individuals' SD in US-expectancies during extinction and CS+/CS- discrimination in the US-expectancies on the first test trial. This result suggests that higher levels of variability during extinction are associated with higher

discriminative fear responding at test. The degree of variability in US-expectancy ratings during extinction was not correlated with differential SCR on the first test trial.

Conclusion and Discussion

Exposure therapy can start with less fear-eliciting stimuli and gradually progress to more distressing stimuli (i.e., hierarchical approach) or the items of an exposure hierarchy can be progressed in a random order. In the current study a procedure was developed to investigate the effect of hierarchical versus random extinction on fear responding using fear conditioning. On Day 1 participants learned that one stimulus (i.e., CS+) was always followed by an electric shock, whereas another stimulus (i.e., CS-) was never followed by shock. During extinction training on Day 2, participants were presented with the CS+, CS-, and eight morphed stimuli between the CS+ and CS-. In the hierarchical extinction (HE) condition, we aimed to model hierarchical exposure by first presenting participants with the CS-, followed by the morph that was most similar to it, subsequently the most similar to that one, etc. Participants in the random exposure (RE) condition were presented with the same set of stimuli in a random order. On Day 3, we tested for fear responding to the CS+, CS- and a third generalization stimulus (GS). No shocks were administered on Day 2 and 3. In line with the assumptions of the Inhibitory Learning Theory (ILT), we expected higher fear responding at test in the HE compared to the RE condition. In addition, higher expectancy violation, higher physiological arousal and higher variability were expected during extinction in the RE condition compared to the HE condition.

The results revealed no differences between the HE and RE condition in differential fear responding on the first test trial, as measured by US-expectancy ratings and SCR. In addition, we did not find evidence for differences in the course of the test phase between both conditions. As predicted, higher expectancy violation, higher physiological arousal and higher levels of variability in US-expectancy ratings were found throughout extinction in the RE condition compared to the HE condition. However, physiological arousal during extinction was not associated with fear responding during test. Opposite to what was predicted, higher levels of variability in US-expectancy ratings and higher expectancy violation during extinction predicted higher differential fear responding in test.

These findings are not in line with the predictions made by ILT. Although random extinction successfully induced higher levels of expectancy violation and physiological arousal, we did not find the predicted effects on subsequent fear responding. Moreover, the observation that higher levels of variability during extinction predicted higher subsequent fear responding

is opposite to previous findings by Culver et al. (2012), Kircanski et al. (2012), and Brown et al. (2017). They found that higher variability throughout exposure was associated with lower fear responding at follow-up. Whereas Brown et al. (2017), similar to the current study, used variability in US-expectancies to test this hypothesis[2], Culver et al. (2012) and Kircanski et al. (2012) focused on variability in *subjective fear ratings*. Arguably, as compared to subjective fear ratings, US-expectancy ratings allow for a more direct measure of the expectancy violation that ILT deems important. One can indeed assume that US-expectancies during extinction reflect expectancy violation given that the US was not administered in this phase (e.g., Sevenster, Beckers, & Kindt, 2013). However, to gauge expectancy violation even more directly, it is recommended for future research to ask participants after each US omission during extinction about their subjective feeling of 'surprise'.

The predictions about random versus hierarchical extinction made by ILT are opposite to what recent work on updating of fear memories predicts (e.g., Gershman et al., 2013). In this line of research, it is assumed that large expectancy violation results in the formation of a new competing (inhibitory) memory, because the abrupt transition from acquisition to extinction indicates a novel *state* prompting the formation of such new memory (e.g., Gershman, Blei, & Niv, 2010). This approach states that if the transition between acquisition and extinction occurs more gradually, expectancy violation (or prediction error) is kept low and instead of the formation of a new memory, new information can be integrated in the original fear memory. Gershman and colleagues (2013) tested this hypothesis in rodents by manipulating the reinforcement schedule to keep the transition between acquisition and extinction gradual and expectancy violation low. They found that gradually rather than abruptly reducing the number of USs during extinction prevented the recovery of fear. In the current study, the transition between acquisition and extinction was made gradual in the HE condition, since we first presented the CS-, which evoked little expectancy violation because it was never paired by shock. Subsequently the morph most similar to the CS- was presented, followed by the morph most similar to that stimulus, etc. This way, the differences between the subsequent stimuli were very small and we moved to the CS+ in a very gradual way, arguably keeping expectancy violation low. However, our findings are not in line with this approach either: hierarchical extinction did not result in less fear responding during test than random extinction. Notably, we found overall lower expectancy violation in the HE condition compared to the RE condition,

[2] Notably, variability during extinction was operationalized by calculating the sum of the absolute value of differences in US-expectancies on each consecutive trial. Similar analyses on our data did, however, not change our conclusions.

but nevertheless US-expectancies increased in the HE condition by the end of extinction training, when being presented with stimuli that were more similar to the CS+ and the CS+ itself.

In theoretical approaches such as ILT, random exposure is seen as one of several possible strategies to induce variability throughout exposure (Craske et al., 2008). The effects of random exposure have typically been tested in combination with other strategies to induce variability, such as the use of multiple stimuli and contexts or varying the timing of exposure trials. So far, two clinical studies directly compared the effects of random versus hierarchical exposure therapy (i.e., Kircanski et al., 2012; Lang & Craske, 2000; see introduction). Despite the use of multiple strategies to induce variability, both studies found no evidence for random exposure being more effective than hierarchical exposure. These results are in line with the results of the current study. It can, however, be considered a unique strength of the current study that both conditions only differed in the order in which the feared stimuli were presented, while controlling for all other variables. These findings imply that other factors might come into play when deciding to choose one approach or the other, such as dropout rates. In the study of Lang and Craske (2000) higher peak fear levels were found in the random group, suggesting that random exposure might be less feasible for clients and associated with higher dropout from treatment. The current study does not provide an answer to this question, but it might be an interesting topic for future research.

In conclusion, we developed a novel procedure to investigate the effects of hierarchical versus random extinction in the laboratory, using a set of morphed stimuli during extinction. Arguably, this procedure is a better model for clinical exposure therapy than regular extinction training to the CS+. Clinical exposure is typically not limited to the CS+ and often the CS+ is unavailable or unidentified. The observation that presenting the stimuli in a random order resulted in higher expectancies, higher physiological arousal and higher variability than presenting them in a hierarchical order, adds to the validity of this procedure. However, the results of the current study do not provide evidence for the superiority of random extinction on long-term fear reduction.

References

Brown, L. A., LeBeau, R. T., Chat, K. Y., & Craske, M. G. (2017). Associative learning versus fear habituation as predictors of long-term extinction retention. *Cognition & Emotion, 31*, 687-698. doi:10.1080/02699931.2016.1158695

Cain, C. K., Blouin, A. M., & Barad, M. (2004). Adrenergic transmission facilitates extinction of conditional fear in mice. *Learning and Memory, 11*, 179-187. doi:10.1101/lm.71504

Choy, Y., Fyer, A. J., & Lipsitz, J. D. (2007). Treatment of specific phobia in adults. *Clinical Psychology Review, 27*, 266-286. doi:10.1016/j.cpr.2006.10.002

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, N., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy, 46*, 5-27. doi: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. doi:10.1016/j.brat.2014.04.006

Culver, N., Stoyanova, M. S., & Craske, M. G. (2012). Emotional variability and sustained arousal during exposure. *Journal of Behavior Therapy and Experimental Psychiatry, 42*, 787-793. doi:10.1016/j.jbtep.2011.10.009

Dawson, M. E., Schell, A. M., & Fillion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G., Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 200-224). Cambridge: Cambridge University Press.

Foa, E. B., Huppert, J. D., & Cahill, S. P. (2006). Emotional processing theory: An update. In B. O. Rothbaum (Ed.), *Pathological anxiety: Emotional processing in etiology and treatment* (pp. 3-24). New York: Guilford Press.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin, 99*, 20-35.

Foa, E. B., & McLean, C. P. (2016). The efficacy of exposure therapy for anxiety-related disorders and its underlying mechanisms: The case of OCD and PTSD. *Annual Review of Clinical Psychology, 12*, 1-28. doi:10.1146/annurev-clinpsy-021815-093533

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review, 117*, 197-209. doi:10.1037/a0017808

Gershman, S. J., Jones, C. E., Norman, K. A., Monfils, M-H., Niv, Y. (2013). Gradual extinction prevents the return of fear: Implications for the discovery of state. *Frontiers in Behavioral Neuroscience, 7*, 164. doi:103389/fnbeh.2013.00164

Hovland, C. I. (1937). The generalization of conditioned responses. III. Extinction, spontaneous recovery, and disinhibition of conditioned and generalized responses. *Journal of Experimental Psychology, 21*, 47-62. doi:10.1037/h0055714

Kircanski, K., Mortazavi, A., Castriotta, N., Baker, A. S., Mystkowski, J. L., Yi, R., & Craske, M. G. (2012). Challenges to the traditional exposure paradigm: Variability in exposure therapy for contamination fears. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 745-751. doi:10.1016/j.jbtep.2011.10.010

Knowles, K. A., & Olatunji, B. O. (2018). Enhancing inhibitory learning: The utility of variability in exposure. *Cognitive and Behavioral Practice*. doi:10.1016/j.cbpra.2017.12.001

Lang, A. J., & Craske, M. G. (2000). Manipulations of exposure-based therapy to reduce return of fear: A replication. *Behaviour Research and Therapy, 38*, 1-12. doi:10.1016/S0005-7967(99)00031-5

Leer, A., Engelhard, I. M., Lenaert, B., Struyf, D., Vervliet, B., & Hermans, D. (2017). Eye movement during recall reduces objective memory performance: An extended replication. *Behaviour Research and Therapy, 92*, 94-105. doi:10.1016/j.brat.2017.03.002

Lenaert, B., Claes, S., Raes, F., Boddez, Y., Joos, E., Vervliet, B., & Hermans, D. (2012). Generalization of conditioned responding: Effects of autobiographical memory specificity. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 60-66. doi:10.1016/j.btep.2010.12.010

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: retrospective revaluation induced by experience and instruction. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*, 97-106. doi:10.1037/0278-7393.29.1.97

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology, 8*, 656-672.

Myers, K. M., & Davis, M. (2007). Mechanisms of fear extinction. *Molecular Psychiatry, 12*, 120-150. doi:10.1038/sj.mp.4001939

Pappens, M., Schroijen, M., Van den Bergh, O., & Van Diest, I. (2015). Retention of perceptual generalization of fear extinction. *International Journal of Psychophysiology, 98*, 520-528. doi:10.1016/j.ijpsycho.2015.01.007

Pineles, S. L., Orr, M. R., & Orr, S. P. (2009). An alternative scoring method for skin conductance responding in a differential fear condition paradigm with a long-duration conditioned stimulus. *Psychophysiology, 46*, 984-995. doi:10.1111/j.1469

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

Sánchez-Meca, J., Rosa-Alcázar, A. I., Marín-Martínez, F., & Gómez-Conesa, A. (2010). Psychological treatment of panic disorder with or without agoraphobia: A meta-analysis. *Clinical Psychology Review, 30*, 37-50. doi:10.1016/j.cpr.2009.08.011

Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction error governs pharmacologically induced amnesia for learned fear. *Science, 339*, 830-833.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*, 36-45. doi:10.1027/1618-3169/a000005

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy, 43*, 357-371. doi:10.1016/j.brat.2004.02.005

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. doi:10.1016/j.cpr.2017.10.010

Chapter 4

Virtually unexpected: No role of expectancy violation in virtual reality exposure for public speaking anxiety

Based on:

Scheveneels, S., Boddez, Y., Van Daele, T., & Hermans, D. (2018). *Virtually unexpected: No role of expectancy violation in virtual reality exposure for public speaking anxiety*. Manuscript submitted for publication.

Abstract

In the current study, we examined the role of expectancy violation in virtual reality exposure therapy (VRET), a key mechanism in exposure therapy according to the inhibitory learning model. Participants with public speaking anxiety were asked to give speeches in virtual reality. We experimentally manipulated whether expectancies related to the overt reactions of the audience could be violated. In the interactive condition, participants received information that the virtual audience could have reacted on the presentation. Participants in the non-interactive condition were told that such interaction was not possible. Additionally, we asked each participant individually to report their expectancies in public speaking situations and which of these could be tested in VRET.

Results show a reduction in public speaking anxiety from pre to post VRET. Treatment outcome was, however, not influenced by the experimental manipulation regarding whether or not the audience was interactive. In addition, correlational analyses revealed that the individually reported proportions of testable expectancies were not predictive for treatment outcome. In conclusion, the results of the current study suggest that the effects of VRET are not univocally explained by the mechanism of expectancy violation. Further research is warranted to identify the exact mechanisms at work in VRET.

Keywords: Virtual reality; Exposure therapy; Expectancy violation; Anxiety

Virtually unexpected: No role of expectancy violation in virtual reality exposure for public speaking anxiety

Exposure involves the repeated confrontation with fear-evoking stimuli or situations and is a key component in the treatment of anxiety (e.g., Öst, Havnen, Hansen, & Kvale, 2015, Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008). Typically, this confrontation takes place in real life. For example, a client with fear of flying would be encouraged to take a flight. However, such exposure in vivo could be demanding for both the client and the therapist. Immediate and direct confrontation with the phobic situation is sometimes considered as too threatening by the client, leading to low treatment acceptance and dropout (e.g., Choy, Fyer, & Lipsitz, 2007). For the therapist, organizing an in vivo exposure session can be time-consuming, for example when one has to gather an audience for an exercise in public speaking anxiety.

Virtual reality exposure therapy (VRET) can overcome some of these challenges. In VRET the phobic situation is generated by a computer using virtual reality technology rather than by the natural environment (e.g., Stever, 1992). This can provide easy access to a wide range of exposure exercises without leaving the therapist office. Moreover, it has been demonstrated that VRET is associated with higher treatment acceptability in phobic individuals (Garcia-Palacios, Botella, Hoffman, & Fabregat, 2007, but see Opriş et al., 2012). Besides these advantages, empirical evidence confirms the effectiveness of VRET for anxiety reduction. Meta-analyses conclude that VRET is superior to waitlist control and that VRET-gains generalize to real life (Morina, Ijntema, Meyerbröker, & Emmelkamp, 2015; Opriş et al., 2012; Powers & Emmelkamp, 2008). Moreover, VRET has found to be equally effective as exposure in vivo on the short- as well as the long-term (Opriş et al., 2012; Powers & Emmelkamp, 2008).[1]

Inhibitory Learning Theory (ILT) explains the success of exposure therapy by the formation of a safe (inhibitory) association (‘taking a plane does not go together with crashing’), which competes with the original (excitatory) fear association (‘taking a plane goes together with crashing’) (Craske et al., 2008; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). Since it is assumed that the relative strength of the inhibitory and excitatory association determines fear responding, ILT postulates that inhibitory learning should be maximized during exposure therapy. In doing so, it is assumed that the concept of expectancy violation plays a central role. In particular, exposure should be directed at providing a strong mismatch between

[1] Notably, systematic evaluation of the research quality of VRET studies reveals generally low research quality (McCann et al., 2014; Motraghi, Seim, Meyer, & Morissette, 2014; Page & Coxon, 2016).

the client's expectancies for the likelihood of an aversive outcome and the actual outcome. The more the expectancy can be violated during exposure, the stronger the inhibitory learning (e.g., Craske et al., 2014; Deacon et al., 2013).

Arguably, the effectiveness of VRET is difficult to explain by the mechanism of expectancy violation. In particular, VRET might be less appropriate to test and violate particular expectancies compared to exposure in vivo because certain outcomes cannot occur in VR. For example, in VRET for fear of flying, feared outcomes like an actual injury, or death because of a plane crash cannot occur. The same holds for animal fears such as dog phobia: the expectancy that the dog would attack and you would sustain injuries from its bites, cannot be tested in VRET. In light of this, the success of VRET seems quite paradoxical.

However, this paradox can be solved in at least two ways. First, a sense of presence, or the connection an individual feels with the VR environment, can explain why an individual experiences the VR environment 'as if it is real' and might come to expect certain outcomes to occur even though it is actually impossible (e.g., Price & Anderson, 2007; Price, Mehta, Tone, & Anderson, 2011). Nevertheless, even then there is the risk that the individual retrospectively reasons that certain aversive outcomes could not occur because treatment took place in VR. Experimental research has shown that such retrospective reasoning about the absence of the aversive outcome can cause a return of fear responding. Using a differential fear conditioning procedure, Raes, De Houwer, Verschuere, and De Raedt (2011) first paired one stimulus with an electric shock. In a subsequent extinction phase, the stimulus was no longer paired with the shock, as an analogue for clinical exposure therapy. After extinction, the experimental group received information that due to a technical failure the electric shock could not occur during the extinction phase. These participants showed higher fear responding in a subsequent test phase compared to a control group that did not receive this information. Translating these results to VRET, retrospective reasoning that the feared outcome could not have occurred might cause relapse.

A second way to solve the paradox is to consider that the occurrence of a limited set of feared outcomes can be tested in VRET. This is particularly true for feared outcomes towards one's own reactions. In the fear of flying example, a plane crash cannot occur in VRET, but expected outcomes related to own reactions such as having a panic attack can. In line with the theoretical assumptions of ILT this would imply that those clients whose fear of flying is primarily driven by the expectancy that the plane will crash might benefit less from VRET

compared to those clients whose fear is driven by the expectancy of having a panic attack during a flight.

In the present study, we examined the role of expectancy violation in VRET in both an experimental and a correlational way. We used a clinical analog sample of participants highly fearful of public speaking. We experimentally manipulated whether expectancies concerning the reactions of the virtual audience could be violated. After two sessions of VRET, participants in the 'interactive condition' were instructed that the virtual audience could have reacted on their presentations. Participants in the 'non-interactive condition' received the information that such interaction was not possible and as such expectations about the reactions of the audience could not be tested and violated in this group. In line with the role of expectancy violation in ILT and the findings of Raes et al. (2011), we hypothesized that VRET would be less effective in the latter group, with higher post-treatment fear responding in a behavioral avoidance task and on self-report questionnaires in this group compared to the former group.

In addition, each participant was asked about his or her expectancies in public speaking situations. After VRET, we examined which of their expectancies participants evaluated as being testable in VRET and calculated the proportion of testable expectancies for each participant. We tested whether the proportion of expectancies that could be tested in VRET was predictive for the outcome of VRET. Moreover, each of the expectancies was categorized as being related to (1) participants' own reactions, (2) the (overt) reactions of the audience, or (3) negative evaluation. It was hypothesized that expectancies with regard to participants' own reactions in a public speaking situation (e.g., having a panic attack, going out of my mind) would be evaluated as being more testable in VRET than expectancies related to the reactions of the audience or being negatively evaluated. Moreover, it was predicted that participants who reported a higher proportion of expectancies related to their own reactions relative to other expectancies would benefit more from VRET.

Method

Participants

Forty-three participants (38 females, 5 males), fearful of public speaking, were randomly allocated to the interactive (n = 22) and non-interactive (n = 21) condition. The mean age of the participants was 22.72 (SD = 4.64). Participants were screened using a two-item questionnaire previously used in studies recruiting participants of public speaking (Culver, Stoyanova, &

Craske, 2012; Niles, Craske, Lieberman, & Hur, 2015; Tsao & Craske, 2000). The items assessed (1) how anxious they would feel when giving a formal speech in front of a live audience and (2) the extent to which they would avoid taking a class that required giving an oral presentation. Answers were given on a scale ranging from 0 = none/never to 8 = extremely/always. Respondents scoring a 6 or higher on anxiety and a 5 or higher on avoidance were recruited for participation. Participants received course credit or financial compensation for their participation. The study was approved by the ethical committee of the Faculty of Psychology and Educational Sciences of the University of Leuven.

Measures and apparatus

Self-report questionnaires

Personal Report of Confidence as a Speaker (PRCS; Paul, 1966). The PRCS consists of 30 true/false items and was used to assess participants' confidence as a public speaker. Previous research supports the internal consistency (r = .91) and validity of the PRCS showing correlations between .52 and .97 with other measures of social anxiety (Daly, 1978; Klorman, Weerts, Hastings, Melamed, & Lang, 1974). Moreover, the PRCS has shown to be sensitive to change after exposure-based treatment (e.g., Lawm, Schwartz, Houlihan, & Cassisi, 1994).

Self-Statements during Public Speaking Scale (SSPS; Hofmann & DiBartolo, 2000). The SSPS is a 10-item questionnaire that is designed to assess what the individual thinks and feels during public speaking. Items are scored on a scale between 0 (= totally disagree) and 5 (= totally agree). Five items measure positive statements (SSPS-P) and five items measure negative statements (SSPS-N). Internal consistency has shown to be high for both the SSPS-P subscale (α = .80) and the SSPS-N subscale (α = .86) (Hofmann & DiBartolo, 2000). Test-retest reliability was acceptable for the SSPS-P subscale (r = .78) as well as the SSPS-N subscale (r = .80). In addition, the SSPS-N subscale was found to be sensitive to treatment, but no significant increase from pre- to post-treatment has been observed in the SSPS-P subscale (Hofmann & DiBartolo, 2000).

List of expectancies. Based on the existing literature and available questionnaires, we developed a list of 50 expectancies commonly reported by individuals with speaking anxiety (see Appendix A). This list contains 25 expectancies about individuals' own reactions and behavior (e.g., I will stutter; I will panic), 14 expectancies related to being negatively evaluated

by others (e.g., They will think that I am incompetent, I will make a bad impression) and 11 expectancies about the overt reactions of the audience (e.g., People in the audience will ask difficult questions, People in the audience will criticize me). Participants rated the expectancies on a yes/no answer format. Before VRET we asked participants to indicate for each listed expectancy whether or not it is applicable to them in public speaking situations. After VRET participants were asked to indicate for each of the expectancies whether or not it was possible to test the expectancy in the VRET exercises (irrespective of whether or not it happened). For each participant, the proportion of testable expectancies was computed by taking the overlap between expectancies the participant reported in public speaking situations (at pre-assessment) and which expectancies could be tested in VRET. This number was then divided by the total number of expectancies reported by the participant in public speaking situations (i.e., at pre-assessment). The proportion of testable expectancies was computed for the total set of expectancies, as well as for each subtype (self, negative evaluation, reactions of audience).

Behavioral Avoidance Test (BAT). In the BAT, participants gave a speech in front of a live audience of 2 females and 1 male. Participants were instructed to speak for as long as they could, up to 2 minutes. After 2 minutes participants were told that they had successfully completed the 2-minute speech and that they had the possibility to continue with their speech for a maximum of another 2 minutes. It was emphasized that this was not mandatory and that it was entirely the choice of the participant whether or not to continue. The duration of the speeches was registered as a behavioral index. At each of the two assessment sessions, a (at the time of the study in Belgium controversial) topic for the BAT speech was randomly picked by the participant out of two possibilities: (1) “Is it desirable to put a ban on headscarves?” and (2) “Should there be limitations on the earnings of managers, sportsmen, etc.?”. Once picked, a topic was removed from the pool. Hence all participants in the end presented about the same topics, but whether a particular topic was picked at pre- or post-assessment could vary. Participants were not allowed to prepare their BAT speeches.

During the BAT, we monitored Subjective Units of Distress (SUDS; Wolpe, 1973). Participants were instructed to give a rating between 0 and 100, where 0 = no fear, 25 = mild fear, 50 = moderate fear, 75 = severe fear, and 100 = very severe fear. SUDS ratings were asked just before the start of the speech, after 1 minute and after 2 minutes, just before participants ended their speech. For each BAT, an average across all SUDS ratings was calculated.

A Polar RS800CX (Polar Electro Oy, Kempele, Finland) was used to measure heart rate (beats per minute) during the BAT. Participants wore a wristband that receives data from a chest-strap also worn by the participant. Inter-beat intervals were sampled at a frequency of 1000 Hz. Empirical evidence supports the validity and reliability of this type of heart rate monitors in research (Weippert et al., 2010; but see Quintana, Heathers, & Kemp, 2012; Wallén, Hasson, Theorell, Canlon, & Osika, 2012). A 5-minute baseline heart rate was measured at pre- and post-assessment while participants were seated and before they received instructions about the BAT. ARTiiFACT software (Kaufmann, Sütterlin, Schulz, & Vögele, 2011) was used for automated artifact detection and for the handling of missing data and deletion of artifacts. After processing, a difference score was computed between the average heart rate measured during the BAT and the average baseline heart rate.

Procedure

The study consisted of four sessions that took place on four consecutive days for any participant enrolling in the study: a pre-assessment session, two VRET sessions and a post-assessment session.

At the start of the pre-assessment session, participants signed the informed consent. Next, baseline heart rate was recorded while participants were left alone in the room and were instructed to sit quietly and remain still. Subsequently, the experimenter guided participants to another room where they received instructions about the BAT. After the BAT, participants completed the PRCS, SSPS and the list with expectancies.

The two VRET sessions were scheduled on the next two days. Each of these sessions contained two exposure exercises. In these exercises participants were asked to give a speech about a (controversial) topic in front of a virtual audience. There were four different topics for the four exercises: (1) “In which circumstances is abortion justified?”, (2) “Are we spending too much time on social media?”, (3) “Is it desirable to put more restrictions on immigration?”, (4) “How should we deal with long-term unemployed?”. All participants presented about every topic, but the order was picked randomly by each participant. During these exercises, participants wore a Samsung Gear VR headset with a Samsung S7 smartphone inserted on which the 360° movie clips were displayed. These movie clips, recorded using a Samsung Gear 360° camera, contained recordings of four audiences varying in composition, duration and context. In the first exercise, the audience consisted of four people seated in an office. The second exercise took place in a meeting room in front of an audience of 12 people. In the third

exercise participants gave their speech in a class room in front of an audience of 20 people. In the fourth exercise an audience of approximately 150 people was seated in an auditorium. Since the movie clips were pre-recorded, they were exactly the same for all participants and audiences were not interactive. Participants had 5 minutes preparation time for each of the speeches. They did not know the duration of the speech or composition of the audience in advance. Before participants started with the first exercise in the first session, they were presented with an empty room to get accustomed to the virtual reality environment and device. At the end of the second VRET session, participants filled in the list of expectancies.

The post-assessment session started with the crucial experimental manipulation. Participants received a sheet with information framed as a non-disclosure agreement with regard to the technology used in the study (see Appendix B). They were instructed to carefully read the information on the sheet and sign it. In the non-interactive condition the information sheet stated that the used technology did not allow for an interactive audience and that therefore the virtual audience could not react on their speeches. It was added that if they, for instance, said something odd or stupid, they would not have noticed this in the reactions (e.g., facial expressions) of the audience. In the interactive condition it was stated that we used new technology that allowed the audience to react on the participants' speeches. Here it was added that if they said something odd or stupid, they could have noticed this in the reactions (e.g., facial expressions) of the audience. This information was repeated orally by the experimenter. Subsequently, 5-minute baseline heart rate was recorded. After this, participants picked a topic and completed the BAT, followed by the PRCS, SSPS and list of expectancies. Finally, we asked participants to indicate how credible it was that the audience was interactive/not interactive. They could choose from four answer possibilities: (1) Not credible at all; (2) Not very credible; (3) Somewhat credible; (4) Very credible.

One week after the post-assessment session, participants completed the PRCS and SSPS online.

Results

Three participants in the non-interactive condition and four participants in the interactive condition rated the instruction about whether or not the audience was interactive as 'not very credible'. All other participants indicated that they found the instruction 'somewhat credible' or 'very credible'. Statistical analyses after exclusion of the participants who rated the

instruction as ‘not very credible’ (N = 36) resulted in the same conclusions as analyses on the entire sample. Here we report the statistical analyses on the entire sample (N = 43).

Is there an effect of the experimental manipulation on VRET outcome?

Repeated measures analyses of variance (rmANOVAs) with Condition (non-interactive, interactive) as between-subjects factor and Time (pre-assessment, post-assessment) as within-subjects factor were conducted to test whether the experimental instruction about whether or not the audience was interactive had an effect on the outcome of VRET.

Manipulation check. As a manipulation check, we compared the proportion of testable expectancies about the overt reactions of the audience before (i.e., after the last VRET exercise) and after the experimental manipulation. We expected a decrease in the proportion of testable expectancies about the (overt) reactions of the audience in the non-interactive condition, but not in the interactive condition. This was expected because participants in the non-interactive condition were informed that the audience could not react on their speeches. This was, however, not confirmed by a 2 (Time: VR, post) × 2 (Condition: interactive, non-interactive) rmANOVA. No significant Time × Condition interaction was found, $F(1, 41) = 0.01$, $p = .98$, $\eta^2_p = .00$. This result suggests that participants’ evaluation of which expectancies about the overt reactions of the audience could be tested during VRET was not influenced by the information that the audience could react to their speeches or not.

BAT. Figure 1 displays the mean SUDS during BAT (left panel) and the mean heart rate during BAT (right panel) for the interactive and non-interactive condition at pre- and post-assessment. The duration of the speeches (behavioral index) was not included in the data analyses since all participants completed the 2-minute BAT speech and only five participants were willing to continue with their speech after the prescribed 2 minutes (preventing reliable analyses).

BAT-SUDS. A 2 (Time: pre, post) × 2 (Condition: interactive, non-interactive) rmANOVA reveals a significant main effect of Time, $F(1, 41) = 52.25$, $p < .001$, $\eta^2_p = .56$. This suggests that VRET was effective in reducing subjective distress during the BAT (see Figure 1). However, no significant Time × Condition interaction was found, $F(1, 41) = 0.74$, p

= .40, η^2_p = .02, indicating that the effect of VRET on SUDS did not differ between the interactive and non-interactive condition.

BAT-HR. A significant main effect of Time was found in a 2 (Time: pre, post) × 2 (Condition: interactive, non-interactive) rmANOVA, $F(1, 41) = 32.15$, $p < .001$, $\eta^2_p = .44$, indicating a decrease in heart rate from pre- to post-assessment BAT. This decrease did not differ between the interactive and non-interactive condition, given a non-significant Time × Condition interaction, $F(1, 41) = 1.06$, $p = .31$, $\eta^2_p = .03$.

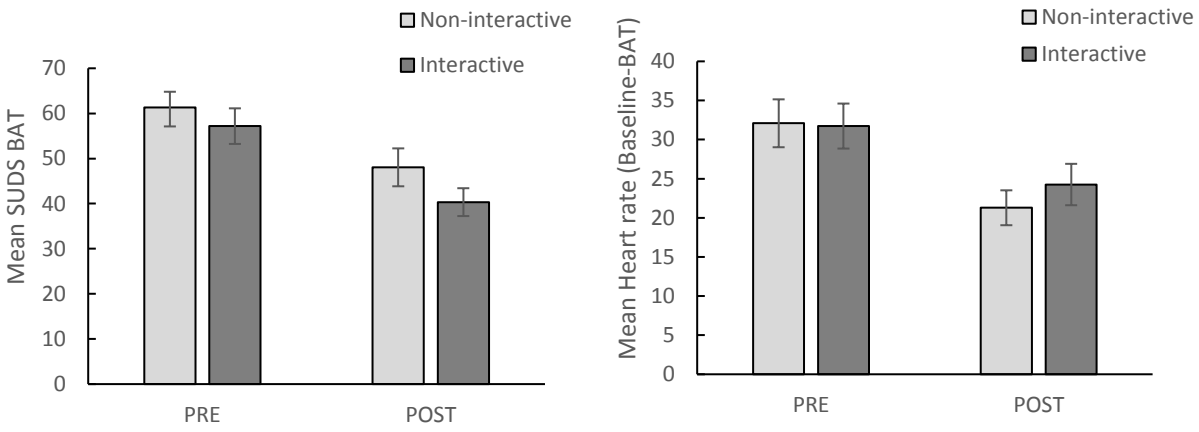

Figure 1. Mean SUDS during BAT (left) and heart rate during BAT (right) per condition at pre-assessment (PRE) and post-assessment (POST). Error bars represent standard errors.

Self-report questionnaires

PRCS. Figure 2 displays the mean PRCS scores per condition at pre-assessment, post-assessment and follow-up. A 2 (Time: pre, post) × 2 (Condition: interactive, non-interactive) rmANOVA reveals a significant main effect of Time, $F(1, 41) = 39.77$, $p < .001$, $\eta^2_p = .49$, but no significant Time × Condition interaction, $F(1, 41) = 0.04$, $p = .84$, $\eta^2_p = .01$. These results demonstrate a reduction in PRCS scores from pre to post VRET with, however, no differences between both conditions. Although Figure 2 suggests an increase in PRCS score from post-assessment to follow-up, a main effect of Time in a 2 (Time: pre, FU) × 2 (Condition: interactive, non-interactive) rmANOVA demonstrated that the results of VRET were (partially) maintained at 1-week follow-up, $F(1, 41) = 15.83$, $p < .001$, $\eta^2_p = .28$. No Time × Condition interaction was found, $F(1, 41) = 1.34$, $p = .25$, $\eta^2_p = .03$.

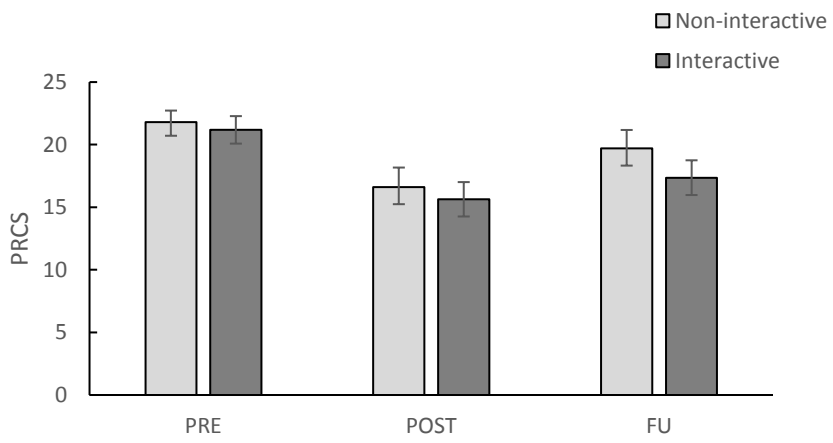

Figure 2. Mean PRCS scores per condition at pre-assessment (PRE), post-assessment (POST) and follow-up (FU). Error bars represent standard errors.

SSPS-P and SSPS-N. Figure 3 displays the mean SSPS-P scores (left panel) and mean SSPS-N scores (right panel) per condition at pre-assessment, post-assessment and follow-up. For the SSPS-P data a significant main effect of Time, $F(1, 41) = 36.79$, $p < .001$, $\eta^2_p = .47$, but no significant Time × Condition interaction, $F(1, 41) = 2.56$, $p = .12$, $\eta^2_p = .06$, was observed in a 2 (Time: pre, post) × 2 (Condition: interactive, non-interactive) rmANOVA. Hence, the effects of VRET were confirmed by this outcome measure, with again no effect of the experimental manipulation. Notably, the effects of VRET on the SSPS-P were not maintained at 1-week follow-up: no significant main effect of Time in a 2 (Time: pre, FU) × 2 (Condition: interactive, non-interactive) rmANOVA $F(1, 41) = 0.23$, $p = .64$, $\eta^2_p = .01$ was found. The Time × Condition interaction was not significant, $F(1, 41) = 0.10$, $p = .76$, $\eta^2_p = .01$.

Similar results were found for the SSPS-N, with a main effect of Time, $F(1, 41) = 36.79$, $p < .001$, $\eta^2_p = .47$, and no Time × Condition interaction, $F(1, 41) = 2.56$, $p = .12$, $\eta^2_p = .06$, in a 2 (Time: pre, post) × 2 (Condition: interactive, non-interactive) rmANOVA. Here, however, effects were maintained at 1-week follow-up as demonstrated by a significant main effect of Time in a 2 (Time: pre, FU) × 2 (Condition: interactive, non-interactive) rmANOVA $F(1, 41) = 6.55$, $p = .01$, $\eta^2_p = .14$. The Time × Condition interaction was not significant, $F(1, 41) = 2.85$, $p = .10$, $\eta^2_p = .07$.

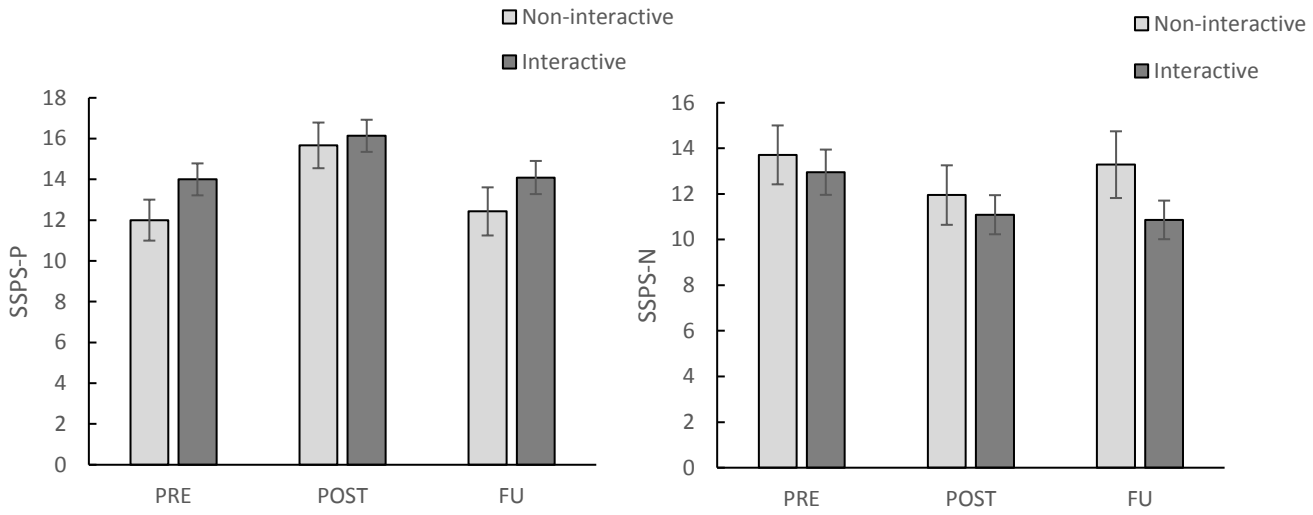

Figure 3. Mean SSPS-P scores (left) and SSPS-N scores (right) per condition at pre-assessment (PRE), post-assessment (POST) and follow-up (FU). Error bars represent standard errors.

Does the proportion of testable expectancies predict VRET outcome?

Using a sum score of the different types of expectancies, we tested whether the individually calculated proportion of testable expectancies (measured before the experimental manipulation) predicted VRET outcome. For each outcome measure (i.e., BAT-SUDS, BAT-HR, PRCS, SSPS-P, SSPS-N) we conducted a linear regression analysis with pre-assessment levels and proportion of testable expectancies as predictor variables and post-assessment levels as the outcome variable. Results of these analyses are displayed in Table 1. Pre-assessment levels significantly predicted post-assessment levels across all outcome measures. However, for none of the outcome measures, the proportion of testable expectancies predicted VRET outcome when controlled for pre-assessment levels.

Table 1

Multiple Linear Regression Analyses Predicting Post Levels of VRET Outcomes From Pre Levels and Proportion of Testable Expectancies

	B	*SE B*	*β*	*t-value*	*p-value*
BAT-SUDS					
Intercept	-4.50	8.56		$t(2) = -0.53$	$p = .60$
Pre level	0.71	0.11	.71	$t(2) = 6.32$	$p < .001$
Expectancies	11.89	7.67	.17	$t(2) = 1.55$	$p = .13$
BAT-HR					
Intercept	7.16	5.20		$t(2) = 1.38$	$p = .18$
Pre level	0.54	0.10	.64	$t(2) = 5.19$	$p < .001$
Expectancies	-2.58	5.55	-0.06	$t(2) = -0.47$	$p = .65$
PRCS					
Intercept	-2.67	4.71		$t(2) = -0.57$	$p = .57$
Pre level	0.85	0.19	.59	$t(2) = 4.53$	$p < .001$
Expectancies	1.14	3.44	.04	$t(2) = 0.33$	$p = .74$
SSPS-P					
Intercept	5.48	1.75		$t(2) = 3.13$	$p = .003$
Pre level	0.75	0.11	.72	$t(2) = 6.57$	$p < .001$
Expectancies	2.27	1.90	.07	$t(2) = 0.67$	$p = .51$
SSPS-N					
Intercept	-0.73	1.30		$t(2) = -0.56$	$p = .58$
Pre level	0.84	0.07	.88	$t(2) = 11.75$	$p < .001$
Expectancies	1.98	1.48	.10	$t(2) = 1.33$	$p = .19$

Are expectancies about individuals' own reactions better testable in VRET?

Figure 3 displays the mean proportions of testable expectancies per Type of expectancy (self, audience, negative evaluation) as measured after VRET (before the experimental manipulation). We tested whether VRET is more eligible to test and violate expectancies about individuals' own reactions compared to expectancies about the overt reactions of the audience or expectancies about being negatively evaluated. This was confirmed by a significant main effect of Type of expectancy in a rmANOVA on the proportion of testable expectancies with Type of expectancy as within-subjects factor, $F(2, 84) = 7.02$, $p = .002$, $\eta^2_p = .14$.

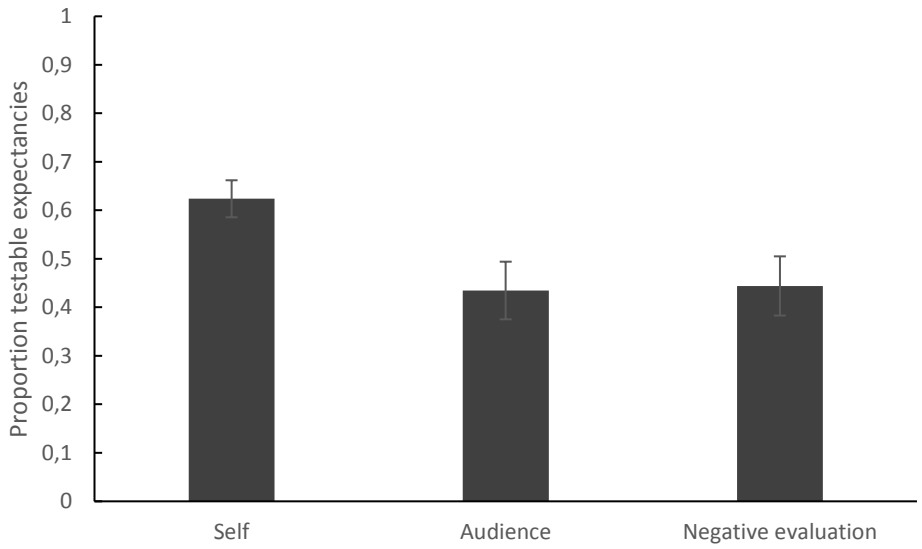

Figure 4. Mean proportions testable expectancies per Type of expectancy (self, audience, negative evaluation) measured after VRET (before the experimental manipulation).

Do individuals with higher proportions of expectancies about their own reactions benefit more from VRET?

Given that expectancies about one's own reactions are better testable in VRET, it can be predicted that individuals with relatively higher proportions of such expectancies benefit more from VRET. To answer this question, we calculated a new variable representing the proportion of expectancies about one's own reactions. This was done by dividing for each participant the number of expectancies about one's own reactions by the total number of expectancies at pre-assessment. Multiple regression analyses were conducted for each outcome measure (i.e., BAT-

SUDS, BAT-HR, PRCS, SSPS-P, SSPS-N) with pre-assessment level and proportion expectancies related to own reactions as predictor variables and post-assessment level as outcome variable. Results are displayed in Table 2. In all outcome measures, pre-assessment levels significantly predicted post-assessment levels. Only in the PRCS data, the proportion of expectancies about one's own reactions significantly predicted post-assessment levels above pre-assessment levels (Table 2). However, the direction of this relation was opposite to what we predicted: higher proportions of expectancies about one's own reactions were related to worse treatment outcome.

Table 2

Multiple Linear Regression Analyses Predicting Post Levels of VRET Outcomes From Pre Levels and Proportion of Expectancies About Own Reaction At Pre-Assessment

	B	*SE B*	*B*	*t-value*	*p-value*
BAT-SUDS					
Intercept	-6.46	10.68		$t(2) = -0.60$	$p = .55$
Pre level	0.69	0.11	.69	$t(2) = 6.13$	$p < .001$
Expectancies	21.33	17.53	.14	$t(2) = 1.22$	$p = .23$
BAT-HR					
Intercept	5.29	6.19		$t(2) = 0.85$	$p = .40$
Pre level	0.55	0.10	.65	$t(2) = 5.40$	$p < .001$
Expectancies	0.17	12.42	0.01	$t(2) = 0.01$	$p = .99$
PRCS					
Intercept	-13.23	5.56		$t(2) = -2.38$	$p = .02$
Pre level	0.94	0.17	.65	$t(2) = 5.41$	$p < .001$
Expectancies	20.17	7.32	.33	$t(2) = 2.76$	$p = .009$
SSPS-P					
Intercept	5.09	2.60		$t(2) = 1.96$	$p = .06$
Pre level	0.76	0.11	.73	$t(2) = 6.73$	$p < .001$
Expectancies	1.95	4.32	.05	$t(2) = 0.45$	$p = .65$
SSPS-N					
Intercept	-0.81	1.89		$t(2) = -0.43$	$p = .67$
Pre level	0.84	0.07	.88	$t(2) = 11.61$	$p < .001$
Expectancies	2.50	3.43	.06	$t(2) = 0.73$	$p = .47$

Conclusion and Discussion

The current study investigated the role of expectancy violation in VRET for public speaking anxiety in an experimental and a correlational way. First, we experimentally manipulated whether expectancies about the overt reactions of the audience could be tested and violated by providing participants with verbal information stating that the audience could interact (interactive condition) or could not interact (non-interactive condition). Second, participants were asked to indicate which expectancies they have in public speaking situations and which expectancies they could test in the VRET. It was tested whether the proportion of testable expectancies calculated for each participant individually predicted VRET outcome. In addition, it was hypothesized that expectancies about individuals' own reactions could better be tested in VRET than expectancies about the overt reactions of the audience or about being negatively evaluated by others. We also predicted that participants reporting a higher proportion of expectancies related to own reactions relative to the other types of expectancies would benefit more from VRET.

The results confirm a reduction in public speaking anxiety from pre to post VRET in all outcome measures. The experimental manipulation of expectancy violation did not influence the outcome of VRET, with no differences between the interactive and non-interactive condition. Notably, participants rated the manipulation instruction as credible. However, a manipulation check comparing the proportion of testable expectancies about the overt reactions of the audience before and after the experimental manipulation did not reveal the expected decrease in the non-interactive condition. Although it could therefore be questioned whether participants took the information into account, we took several precautions to guarantee this: participants were told to carefully read the information sheet with experimental instructions, to sign it, and the experimenter also orally repeated the gist of the instruction.

Notably, our results are in line with findings of a study by Morina, Brinkman, Hartanto, and Emmelkamp (2014). The aim of this study was different from the current study, since they investigated whether the degree of interaction in VR had an effect on the sense of presence participants felt. Participants engaged in free speech dialogues with virtual humans and were assigned to either a condition in which interaction between the individual and the virtual human was possible or to a condition in which such interaction was not possible. Hence, instead of giving verbal information, in this study the devices that were used did or did not allow for interaction. Higher levels of presence were reported by the interactive condition. Importantly, despite the fact that participants were not selected based on high (social) anxiety and no test for

generalization to real life was included, there was no effect of the degree of interaction on the reported anxiety levels.

In addition, the correlational findings confirm the conclusion that expectancy violation does not influence the effectiveness of VRET. These findings indeed indicate that the proportion of testable expectancies in VRET did not predict VRET outcome. As hypothesized, expectancies about individuals' own reactions could better be tested in VRET compared to expectancies about the overt reactions of the audience and expectancies about being negatively evaluated. However, the hypothesis that participants with higher proportions of expectancies about their own reactions would benefit more from VRET was not confirmed.

Two remarks can be made with regard to these correlational findings. First, it is possible that participants were not accurate in identifying their expectancies and in evaluating whether these expectancies could be tested during VRET. This could explain why we did not find correlations between (testable) expectancies and VRET outcome. However, ILT assumes that clients are aware of and can report their expectancies, which is a necessary requirement for designing an exposure session in which these expectancies can be maximally violated. For example, before an exposure exercise the client is asked about what he or she expects to happen and afterwards the exercise is evaluated with respect to whether the expected outcome occurred (Craske et al., 2014; Weisman & Rodebaugh, 2018). Second, it is possible that participants report certain expectancies that could be tested in VRET, but that these expectancies are not crucial in their public speaking anxiety. For example, a participant might expect that he will sweat heavily, but might not bother about this. Whether or not this expectancy is then violated might not contribute to the success of the exposure treatment. However, we found significant positive correlations between the number of expectancies and anxiety levels at pre-treatment, which goes against this explanation[2].

The results of the current study suggest that the effects of VRET are not univocally explained by the mechanism of expectancy violation. The question then arises which mechanism(s) are at play in VRET and can account for its effectiveness. According to Emotional Processing Theory (EPT), which has dominated the field for many years, the efficacy of exposure treatment results from the initial activation of fear followed by sustained exposure until fear declines (as originally formulated by Foa and Kozak, 1986). In particular, it is assumed that fear habituation is the crucial mechanism by which new incompatible information

[2] PRCS: $r = .63$ ($p < .001$); BAT_SUDS: $r = .39$ ($p = .01$); SSPS-N: $r = .62$ ($p < .001$).

is available and the activated ‘fear structure’, consisting of propositions between stimuli, responses and meanings in memory, is *replaced* by a ‘non-fear structure’ (Foa & Kozak, 1986; Lang, 1971; Rachman, 1980). In line with the basic assumptions of EPT, it has been demonstrated that VRET allows to evoke fear responding followed by decreases in for instance heart rate and skin conductance (e.g., Wiederhold, Jang, Kim, & Wiederhold, 2002; Wilhelm et al., 2005; but see Mühlberger, Bülthof, Wiedemann, & Pauli, 2007). The current study, however, did not include physiological and verbal measures of fear during VRET and as a consequence does not allow for investigating whether habituation in these measures can predict treatment outcome. This would be an interesting question for future research. However, it should be added that an extensive amount of research shows that habituation is not predictive for the long-term outcome of in vivo exposure therapy (e.g., Baker et al., 2010; Brown, LeBeau, Chat, & Craske, 2017; Culver et al., 2012; Kircanski et al., 2012).

In conclusion, our findings confirm the effectiveness of VRET and suggest that low-cost VR technology could be sufficient for treating anxious individuals. Importantly, the availability of a low-cost device can facilitate the broader dissemination and usage of VRET (Morina et al., 2014). Nevertheless, the exact underlying mechanisms that are at work in VRET are currently not well-known and remain an important topic for further research. Especially since this knowledge can drive the further optimization of VRET.

References

Baker, A., Mystkowski, J., Culver, N., Yi, R., Mortazavi, A., & Craske, M. G. (2010). Does habituation matter? Emotional processing theory and exposure therapy for acrophobia. *Behaviour Research and Therapy, 48,* 1139-1143. doi:10.1016/j.brat.2010.07.009

Brown, L. A., LeBeau, R. T., Chat, K. Y., & Craske, M. G. (2017). Associative learning versus fear habituation as predictors of long-term extinction retention. *Cognition and Emotion, 31,* 687-698. doi:10.1080/02699931.2016.1158695

Choy, Y., Fyer, A. J., & Lipsitz, J. D. (2007). Treatment of specific phobia in adults. *Clinical Psychology Review, 27,* 266-286. doi:10.1016/j.cpr.2006.10.002

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,* 5-27. doi: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. doi:10.1016/j.brat.2014.04.006

Culver, N. C., Stoyanova, M., & Craske, M. G. (2012). Emotional variablility and sustained arousal during exposure. *Journal of Behavior Therapy and Experimental Psychiatry, 43,* 787-793. doi:10.1016/j.btep.2011.10.009

Daly, J. A. (1978). The assessment of social-communicative anxiety via self-reports: A comparison of measures. *Communication Monographs, 45*, 204-218.

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,* 588-596. doi:10.1016/j.brat.2013.06.006

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin, 99,* 20-35.

Garcia-Palacios, A., Botella, C., Hoffman, H., & Fabregat, S. (2007). Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *CyberPsychology & Behavior, 10,* 722-724. doi:10.1089/cpb.2007.9962

Hofmann, S. G., & DiBartolo, P. M. (2000). An instrument to assess self-statements during public speaking: Scale development and preliminary psychometric properties. *Behavior Therapy, 31,* 499-515. doi:10.1016/S0005-7894(00)80027-1

Kircanski, K., Mortazavi, A., Castriotta, N., Baker, A. S., Mystkowski, J. L., Yi, R., & Craske, M. G. (2012). Challenges to the traditional exposure paradigm: Variability in exposure therapy for contamination fears. *Journal of Behavior Therapy and Experimental Psychiatry, 43,* 745-751. doi:10.1016/j.jbtep.2011.10.010

Lang, P. J. (1971). The application of psychophysiological methods to the study of psychotherapy and behavior modification. In A. Bergin & S. Garfield (Eds.), *Handbook of psychotherapy and behavior change.* New York, NY: Wiley.

Kaufmann, T., Sütterlin, S., Schulz, S. M., & Vögele, C. (2011). ARTiiFACT: A tool for heart rate artifact processing and heart rate variability analysis. *Behavior Research Methods, 43,* 1161-1170. doi:10.3758/s13428-011-0107-7

Klorman, R., Weerts, T., C., Hastings, J. E., Melamed, B. G., & Lang, P. J. (1974). Psychometric description of some specific-fear questionnaires. *Behaviour Therapy, 5,* 401-409.

Lawm, G. D., Schwartz, D., Houlihan, D., & Cassisi, J. E. (1994). Graduated exposure plus feedback in the treatment of speech anxiety. *Behavioral Interventions, 9,* 213-223. doi:10.1002/bin.2360090403

McCann, R. A., Armstrong, C. M., Skopp, N. A., Edwards-Stewart, A., Smolenski, D. J., June, J. D., …, Reger, G. M. (2014). Virtual reality exposure therapy for the treatment of anxiety disorders: An evaluation of research quality. *Journal of Anxiety Disorders, 28,* 625-631. doi:10.1016/janxdis.2014.05.010

Morina, N., Brinkman, W., Hartanto, D., & Emmelkamp, P. M. G. (2014). Sense of presence and anxiety during virtual social interactions between a human and virtual humans. *PeerJ, 2,* e337. doi:10.7717/peerj.337

Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. G. (2015). Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy, 74,* 18-24. doi:10.1016/j.brat.2015.08.010

Motraghi, T. E., Seim, R. W., Meyer, E. C., & Morissette, S. B. (2014). Virtual reality exposure therapy for the treatment of posttraumatic stress disorder: A methodological review using CONSORT guidelines. *Journal of Clinical Psychology, 70,* 197-208. doi:10.1002/jclp.22051

Mühlberger, A., Bülthoff, H. H., Wiedemann, G., & Pauli, P. (2007). Virtual reality for the psychophysiological assessment of phobic fear: Responses during virtual driving. *Psychological Assessment, 19,* 340-346.

Niles, A, N., Craske, M. G., Lieberman, M. D., & Hur, C. (2015). Affect labeling enhances exposure effectiveness for public speaking anxiety. *Behaviour Research and Therapy, 68*, 27-36. doi:10.1016/j.brat.2015.03.004

Opriş, D., Pintea, S., García-Palacios, A., Botella, C., Szamosközi,, Ş., & David, D. (2012). Virtual reality exposure therapy in anxiety disorders: A quantitative meta-analysis. *Depression and Anxiety, 29*, 85-93. doi:10.1002.da.20910

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

Page, S., & Coxon, M. (2016). Virtual reality exposure therapy for anxiety disorders: Small samples and no controls. *Frontiers in Psychology, 7*, 326. doi:10.3389/fpsyg.2016.00326

Paul, G. L. (1966). Insight versus desensitization in psychotherapy: An experiment in anxiety reduction. Stanford, CA: Stanford University Press.

Powers, M. B., & Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorder: A meta-analysis. *Journal of Anxiety Disorders, 22*, 561-569. doi:10.1016/j.anxdis.2007.04.006

Price, M., & Anderson, P. (2007). The role of presence in virtual reality exposure therapy. *Journal of Anxiety Disorders, 21*, 742-751. doi:10.1016/j.janxdis.2006.11.002

Price, M., Mehta, N., Tone, E. B., & Anderson, P. L. (2011). Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure for social phobia. *Journal of Anxiety Disorders, 25*, 763-770. doi:10.1016/j.janxdis.2011.03.004

Quintana, D. S., Heathers, J. A. J., & Kemp, A. H. (2012). On the validity of using the Polar RS800 heart rate monitor for heart rate variability research. *European Journal of Applied Physiology, 112*, 4179-4180. doi:10.1007/s00421-012-2453-2

Rachman, S. (1980). Emotional processing. *Behaviour Research and Therapy, 18*, 51-60.

Raes, A. K., De Houwer, J., Verschuere, B., & De Raedt, R. (2011). Return of fear after retrospective inferences about the absence of an unconditioned stimulus during extinction. *Behaviour Research and Therapy, 49*, 212-218. doi:10.1016/j.brat.2010.12.004

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M.

(1995). Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *The American Journal of Psychiatry, 152*, 626-628.

Stever, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communications, 42*, 73-93.

Tsao, J. C. I., & Craske, M. G. (2000). Timing of treatment and return of fear: Effects of massed, uniform-, and expanding-spaced exposure schedules. *Behavior Therapy, 31*, 479-497. doi:10.1016/S0005-7894(00)80026

Wallén, M. B., Hasson, D., Theorell, T., Canlon, B., & Osika, W. (2012). Possibilities and limitation of the polar RS800 in measuring heart rate variability at rest. *European Journal of Applied Physiology, 112*, 1153-1165. doi:10.1007/s00421-011-2079-9

Weippert, M., Kumar, M., Kreuzfeld, S., Arndt, D., Rieger, A., & Stoll, R. (2010). Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto, t6 and an ambulatory ECG system. *European Journal of Applied Physiology, 109*, 779-786. doi:10.1007/s00421-010-1415-9

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. doi:10.1016/j.cpr.2017.10.010

Wiederhold, B. K., Jang, D. P., Kim, S. I., & Wiederhold, M. D. (2002). Physiological monitoring as an objective tool in virtual reality therapy. *CyberPsychology & Behavior, 5*, 77-82. doi:10.1089/109493102753685908

Wiederhold, B. K., & Wiederhold, M. D. (2005). Virtual reality therapy for anxiety disorders: Advances in evaluation and treatment. Washington, DC, US: American Psychological Association. doi:10.1037/10858-000

Wilhelm, F. H., Pfaltz, M. C., Gross, J. J., Mauss, I. B., Kim, S. I., & Wiederhold, B. K. (2005). Mechanisms of virtual reality exposure therapy: The role of the bahavioral activation and behavioral inhibition systems. *Applied Psychophysiology and Biofeedback, 30*, 271-284. doi:10.1007/s10484-005-6383-1

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

Wolpe, J. (1973). The practice of behavior therapy. New York, NY: Pergamon Press.

Appendix A: List of expectancies related to individuals' own reactions (S), the overt reactions of the audience (A) and negative evaluation (NE).

1. I won´t get my facts straight (S)	Yes – No
2. They will think that I am incompetent (NE)	Yes – No
3. I will sweat heavily (S)	Yes – No
4. I will have cardiac palpitations (S)	Yes – No
5. They will think that I am not interesting (NE)	Yes – No
6. I will lose control (S)	Yes – No
7. They will think that I am a weird person (NE)	Yes – No
8. I will start blushing (S)	Yes – No
9. People in the audience will ask difficult questions (A)	Yes – No
10. The anxiety will be intolerable (S)	Yes – No
11. I will not be able to cope with it (S)	Yes – No
12. They will think that I am ridiculous (NE)	Yes – No
13. They will think that I am boring (NE)	Yes – No
14. I will freeze (S)	Yes – No
15. I will escape from the situation (S)	Yes – No
16. People in the audience will sigh in reaction to my speech (A)	Yes – No
17. I will suddenly not know what to say (S)	Yes – No
18. They will think that I am weak (NE)	Yes – No
19. I will vomit (S)	Yes – No
20. People in the audience will have a disapproving look on their face in reaction to my speech (A)	Yes – No
21. I will start crying (S)	Yes – No
22. I won't be able to think (S)	Yes – No
23. People in the audience will yawn in reaction to my speech (A)	Yes – No
24. I will feel dizzy (S)	Yes – No
25. I will hyperventilate (S)	Yes – No
26. People in the audience will make fun of me because of my speech (A)	Yes – No
27. I will be overwhelmed by anxiety and won´t be able to speak anymore (S)	Yes – No
28. People in the audience will criticize me (A)	Yes – No

29. I will feel nauseous (S)	Yes – No
30. They will think that I am unintelligent (NE)	Yes – No
31. I will stutter (S)	Yes – No
32. I will fail (i.e., not succeed in the presentation) (NE)	Yes – No
33. People in the audience will be looking on their mobile phone because they are not interested in my speech (A)	Yes – No
34. I will faint (S)	Yes – No
35. People in the audience will roll their eyes in reaction to my speech (A)	Yes – No
36. People in the audience will frown in reaction to my speech (A)	Yes – No
37. They will think that I am stupid (NE)	Yes – No
38. I will behave hysterically (S)	Yes – No
39. They will think that I am saying stupid things (NE)	Yes – No
40. I will make a bad impression (NE)	Yes – No
41. I will tremble (S)	Yes – No
42. They will think that I look silly (NE)	Yes – No
43. I will panic (S)	Yes – No
44. People in the audience will start talking to each other because they find my speech not interesting (A)	Yes – No
45. I will become crazy (S)	Yes – No
46. They will notice that I am anxious and nervous (NE)	Yes – No
47. I will behave aggressively (S)	Yes – No
48. I will talk in a strange way (S)	Yes – No
49. They will think that I am an idiot/loser (NE)	Yes – No
50. People in the audience will leave in reaction to my speech (A)	Yes – No

Appendix B: Information sheet that served as the experimental manipulation

Interactive condition:

Below, you can find additional information about the virtual environment in which you practiced public speaking:

Despite the fact that it was a virtual audience, the audience was interactive and as a consequence the reactions of the audience were adapted to your presentations. During your presentations, the experimenter could manipulate the reactions of the audience on his PC. This implies that if you, for example, would have said something odd or stupid, you could have noticed this in the reactions (facial expression, laughter, other behavior) of the audience.

This technology (including a photorealistic interactive virtual audience) is relatively new within the field of virtual reality and complex to program.

Because we aim to further optimize this technology and eventually launch it for commercial use, we ask you to declare that you understood the above information and that you will not communicate about the used technology with other research centers or companies involved in the field of virtual reality.

Non-interactive condition:

Below, you can find additional information about the virtual environment in which you practiced public speaking:

Because it was a virtual audience, the audience was not interactive and as a consequence the reactions of the audience were not adapted to your presentations. The virtual environments consisted of 360 degrees movie clips of an audience that were recorded beforehand and were showed during your speeches. This implies that if you, for example, would have said something odd or stupid, you could not have noticed this in the reactions (facial expression, laughter, other behavior) of the audience.

Interactive virtual environments are currently available, but only when avatars are used. In this study, we choose for a photorealistic audience. To date, a photorealistic audience that is able to interact with the user is too complex to program. Therefore we used 360 degrees movie clips of a photorealistic audience that could not interact.

Because these movie clips will be further used for commercial purposes, we ask you to declare that you understood the above information and that you will not communicate about the used technology with other research centers or companies involved in the field of virtual reality.

Chapter 5

Ruining the surprise:
The effect of safety information before extinction on return of fear

Based on:

Scheveneels, S., Boddez, Y., De Ceulaer, T., & Hermans, D. (2019). Ruining the surprise: The effect of safety information before extinction on return of fear. *Journal of Behavior Therapy and Experimental Psychiatry, 63,* 73-78. doi:10.1016/j.jbtep.2018.11.001

Abstract

Background and Objectives. In psychoeducation before exposure treatment patients are sometimes provided with information about the (low) probability that the feared outcome would occur. Since it has been proposed in the literature that this might have adverse effects, the current study investigated the effect of providing participants with this type of safety information on return of fear.

Method. In an ABA-renewal paradigm, participants in the experimental group were instructed between acquisition and extinction that the probability of US-occurrence would be extremely small in the remainder of the experiment. Participants in the control group did not receive this information.

Results. Less return of fear in US-expectancy ratings was observed in participants who received the safety information.

Limitations. We failed to find successful acquisition in the skin-conductance data, which prevented us from interpreting the results of this outcome measure.

Conclusions. These results suggest that providing safety information is not deleterious for the effects of exposure and can even be beneficial for its effects. However, further clinical research is needed.

Keywords: Exposure therapy; Fear extinction; Return of fear; Psychoeducation

Ruining the surprise: The effect of safety information before extinction on return of fear

Exposure therapy has demonstrated its efficacy for a variety of anxiety disorders (e.g., Hofmann & Smits; Norton & Price, 2007). During exposure the patient is encouraged to approach the stimulus that elicits anxiety. Despite its general efficacy, some patients refuse to engage in exposure-based treatments or quit in a later stage of treatment (e.g., Haby, Donelly, Corry, & Vos, 2006). Refusal and dropout rates for exposure-based treatment vary amongst anxiety disorders and range between 20% and 43% for obsessive-compulsive disorder (Foa et al., 2005; Stanley & Turner, 1995; Whittal, Thordarson, & McLean, 2005), 7% and 31% for panic disorder (Cox, Endler, Lee, & Swinson, 1992), 14% and 20% for posttraumatic stress disorder (Hembree et al., 2003; Van Etten & Taylor, 1998), 0% and 45% for specific phobias (Choy, Fyer, & Lipsitz, 2007), and 0% and 27% for social phobia (Feske & Chambless, 1995).

Providing patients with psychoeducation before the start of exposure can potentially reduce refusal and dropout rates by increasing the credibility and acceptability of treatment (e.g., Bluett, Landy, Twohig, & Arch, 2016). Importantly, psychoeducation can consist of different components that can either be provided separately or can be combined with each other, including explaining the treatment rationale (Arch, Twohig, Daecon, Landy, & Bluett, 2015), informing patients about the (physiological) components of anxiety (Norr, Norman, & Schmidt, 2017), and providing objective information about the feared object or situation. The latter component can more specifically target the tendency of anxiety patients to overestimate threat (i.e., the probability that a negative outcome would occur), which can subsequently lower the threshold to engage in a confrontation with the fear-eliciting situation (Vander Haegen & Etienne, 2016)[1]. For example, a patient with fear of flying who is afraid of dying in a plane crash can be provided with objective information about the probability that a plane crashes. The patient might then come to evaluate that it is extremely unlikely to be involved in a plane crash which might facilitate his or her engagement in exposure to taking a plane.

However, a recent approach of exposure therapy with roots in the Inhibitory Learning Theory (ILT; Craske et al., 2008; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014), presumes that there might be disadvantages to providing information about the probability that the feared outcome would occur before the start of exposure. ILT relies heavily on a classical

[1] Notably, this component of psychoeducation shows considerable overlap with cognitive interventions (Clark & Beck, 2010).

fear conditioning framework in which it is assumed that an excitatory association between a conditioned stimulus (CS; e.g., taking a plane) and an unconditioned stimulus (US; e.g., dying in a plane crash) is formed in memory in anxious individuals. During exposure, an additional inhibitory link with the CS is formed (CS-noUS; taking a plane does not go together with dying in a plane crash; Bouton, 1993; Bouton & King, 1983). Since it is assumed that fear responding depends on the relative strength of the excitatory and inhibitory associations, the effects of exposure therapy can be maximized by strengthening inhibitory learning. In strengthening the inhibitory association, the concept of *expectancy violation* plays a crucial role. This refers to the mismatch between the expected outcome and the actual outcome. In line with the Recorla-Wagner model (1972), ILT assumes that more (inhibitory) learning can take place if the mismatch between the expected and experienced outcome is large.

Building on the assumptions of ILT, it has been suggested in the literature that providing patients with information about the (low) probability of the occurrence of a feared outcome, decreases the room for expectancy violation during exposure (Craske et al., 2014). If, in the fear of flying example, the expectancy of a plane crash decreases after having received information about the extremely low odds of plane crashes, taking a plane without crashing would not evoke as much "surprise" or expectancy violation. As such, providing this type of psychoeducation or cognitive interventions before or during the exposure, has been argued to be deleterious to inhibitory learning and to the effectiveness of the exposure treatment (Craske et al., 2014).

In the current study, we investigated the effects of safety information on return of fear using a fear conditioning procedure. In an ABA contextual renewal paradigm, participants first learned in context A that one of the stimuli was always paired with an electric shock (i.e., danger cue; CS+) and the other stimulus was never paired with the shock (i.e., safe cue; CS-). Crucially, before the extinction phase, participants in the experimental group were presented with a verbal instruction about the low probability of US-occurrence. Participants in the control group did not receive this information. Subsequently, extinction took place in context B with both the CS+ and CS- presented in the absence of the electric shock. After extinction, we tested for contextual renewal by presenting the CS+ and CS- in context A again. In line with the assumptions of ILT, it was predicted that the experimental group would show lower shock expectancies and skin conductance responding (SCR) at the start of as well as throughout extinction and higher contextual renewal at test compared to the control group.

Method

Participants

Eighty-two first-year psychology students and community volunteers participated in the experiment in return for payment (8 euro) or course credit. Thirty-six participants were recruited for the control group and 46 for the experimental group. Participants in the experimental group who rated the psychoeducational message as "not believable" or "not very believable" were excluded from the data analysis (Mertens & De Houwer, 2016), leaving a total sample of 72 participants (M_{age} = 20.33; *SD* = 3.04; 57 females) or 36 participants per group. Exclusion criteria were pregnancy, cardio-pulmonary conditions, psychiatric or neurological disorders (e.g., epilepsy) and wrist pain. Before the start of the experiment, all participants gave informed consent. The study was approved by the ethical committee of the Faculty of Psychology and Educational Sciences of the University of Leuven.

Apparatus and Stimuli

Conditioned stimuli and contexts. Two geometrical shapes (i.e., a triangle and a square) were used as conditioned stimuli (CS). The triangle and square had a grey color with a black border and were presented on a 19-inch Dell monitor (type P1911, resolution: 1440 × 900 at 60 Hz). Which of the two geometrical shapes functioned as the CS+ and which one as the CS- was counterbalanced. The background colors of the computer screen served as contexts and were yellow (RGB 255, 255, 128) or blue (RGB 0. 255, 255).

Unconditioned stimuli. A 2 ms electrocutaneous stimulus served as the unconditioned stimulus (US). It was administered to the participant's right wrist by a Digitimer DS7A constant current stimulator (Hertfordshire, UK) through a pair of V91-01 8-mm reusable Bilaney Ag/AgCl electrodes. These electrodes were filled with K-Y jelly.

Measures

US-expectancy ratings. Participants rated their expectancy for the US on an eleven-point scale ranging from 0 = "certainly no shock" to 10 = "certainly shock". They could register their response by a left mouse click on the position of the scale that corresponded to their expectancy.

This was done on a trial-by-trial basis. The rating scale appeared onscreen 200 ms after stimulus onset and remained there for maximum 7 s or until participants gave their response.

Skin conductance response (SCR). A Coulbourn LabLinc V Isolated Skin Conductance coupler (model V71-23, Coulbourn Instruments, Allentown, PA) was used to measure electrodermal responding. This device applied a constant voltage of 0.5 Volts through a pair of disposable Biopac EL 507 electrodes (contact area = 95 mm^2). These electrodes were filled with isotonic paste and attached to the hypothenar site of the left-hand palm. Electrodermal activity was recorded from 2 s prior stimulus onset until 6 s after stimulus offset. The analog signal was digitized at 10 Hz by a NI PCI 3221 data acquisition card (National Instruments Corporation, Austin, Texas).

Procedure

After participants gave informed consent, the electrodes were attached. Using a standard work-up procedure, the intensity of the US was set to a level that was "definitely uncomfortable, but not painful". Before the start of the experimental task, participants were explained that they would be presented with two pictures of geometrical shapes and that one of these shapes could be followed by an electric shock. Participants were told that it was their task to predict the occurrence of the shock and that they could do this by using the rating scale. Subsequently, participants could practice using the rating scale in three practice trials after which they received feedback on whether they used the scale correctly. No CSs or USs were presented during these practice trials.

Table 1 displays a schematic overview of the experimental phases. The experimental task started with one non-reinforced presentation of the CS+ and one CS- presentation to weaken orienting responses in the skin conductance measures (pre-exposure phase). This phase was immediately followed by an acquisition phase that consisted of four presentations of the CS+ that was always followed by the US and four CS- presentations in absence of the US. During acquisition, all stimuli were presented against a blue background of the computer screen (context A).

After acquisition, both groups were presented with the following information: "*You will now continue with the experiment. It is still your task to indicate on the rating scale to what extent you expect that the shock will follow the geometrical shapes*".

Participants in the experimental group received additional safety information. Specifically, this information stated the following: *"At the start of the experiment it was mentioned that one of the geometrical shapes could be followed by the electric shock. For the remainder of the experiment, however, the probability that an electric shock will follow this shape is extremely small (1/1000).*" All information was presented against a white background, similar to the instructions at the start of the experiment.

Subsequently, the CS+ and CS- were presented eight times in the absence of shock during the extinction phase. Notably, during extinction the color of the background screen switched to yellow (context B). Extinction was immediately followed by an ABA-renewal test phase in which the CS+ and CS- were presented three times against the blue acquisition context (context A).

After the experimental task, participants in the experimental group rated to what extent the safety information was believable. They could select one of four options: "not believable", "not very believable", "very believable", and "completely believable".

In order to measure a baseline for the skin conductance, all trials started with a 2 s blank screen. CSs stayed on screen for 8 s and intertrial intervals (ITI) were on average 10 s (range 8-12 s). On US-present trials, the US was delivered 7.5 s after CS onset. For half of the participants the CS+ was presented during the first trial of all experimental phases, for the other half the CS- was presented first. This way we aimed to control for order-effects or the impact of (un)reinforced CS+ presentations on subsequent ratings (Lovibond, 2003; Vervliet, Vansteenwegen, Baeyens, Hermans, & Eelen, 2005). Trial order, stimulus presentation, ITI and registration of the dependent variables were controlled by Affect 4.0 (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010).

Table 1

Overview of the experimental phases

	Pre-exposure	Acquisition		Extinction	Test
Experimental group	CS+ (1) CS- (1)	CS+ (4) CS- (4)	**Safety information**	CS+ (8) CS- (8)	CS+ (4) CS- (4)
Control group	CS+ (1) CS- (1)	CS+ (4) CS- (4)		CS+ (8) CS- (8)	CS+ (4) CS- (4)

Note. CSs are pictures of geometrical shapes (counterbalanced). During acquisition '+' refers to the administration of the US and '-' to the absence of the US. During the other phases no USs are administered. The number of trials is indicated between parentheses. The background coloring refers to the context, operationalized as the background of the computer screen.

Results

Only the results of the US-expectancy ratings are reported, since the skin conductance measure failed to show differential acquisition which is a prerequisite for assessing its extinction and later return (e.g., Boddez, Baeyens, Hermans, & Beckers, 2013; Boddez, Baeyens, Hermans, Van der Oord, & Beckers, 2013).

Figure 1 displays the mean US-expectancy ratings throughout the experiment per CS in the experimental and control groups. A Greenhouse-Geisser correction was applied when the assumption of sphericity was violated.

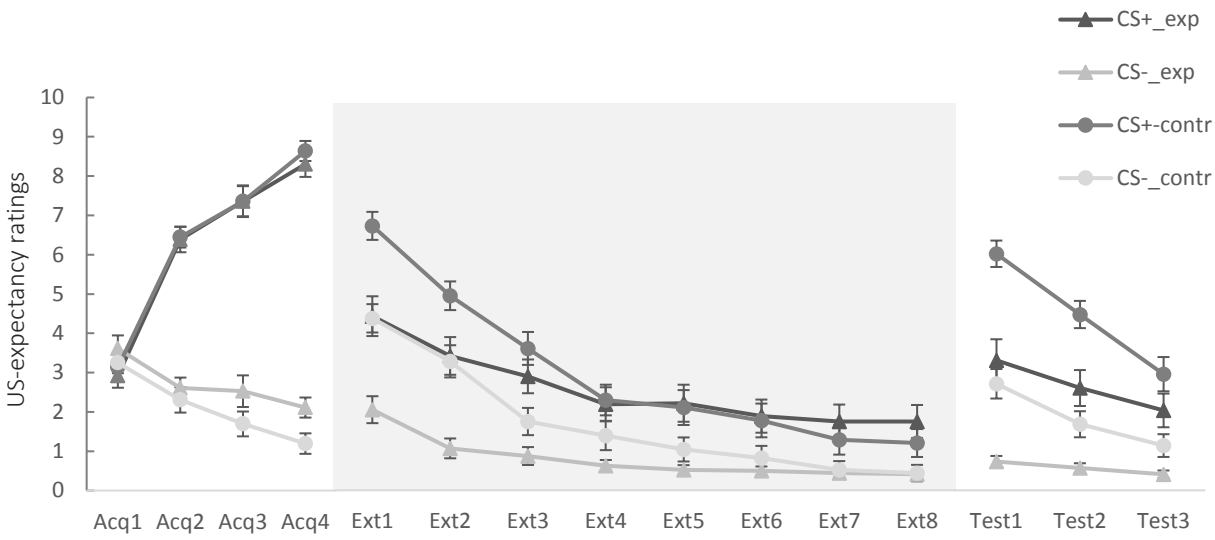

Figure 1. Mean US-expectancy ratings for the four acquisition trials, eight extinction trials and three test trials per CS in the experimental group (exp) and control group (contr). Background colors represent the experimental contexts. Error bars represent standard error of the means.

Acquisition phase

The left panel of Figure 1 suggests in both groups an increase in US-expectancies from the first to the last acquisition trial for the CS+ and a decrease for the CS-. This was confirmed by a 2 (Stimulus) × 2 (Trial) × 2 (Group) repeated measures Analysis of Variance (rmANOVA) comparing the first and last acquisition trial. This analysis revealed a main effect of Stimulus, $F(1, 70) = 173.65$, $p < .001$, $\eta^2_p = 0.71$, a main effect of Trial, $F(1, 70) = 85.51$, $p < .001$, $\eta^2_p = 0.55$, and most importantly a significant Stimulus × Trial interaction, $F(1, 70) = 315.80$, $p < .001$, $\eta^2_p = 0.82$. No effects of Group were found. These results indicate successful acquisition learning which was similar in both groups.

First extinction trial

The middle panel of Figure 1 suggests higher overall US-expectancy ratings in the control group compared to the experimental group on the first extinction trial but no differences between both groups in CS+/CS- discrimination. A 2 (Stimulus) × 2 (Group) rmANOVA on the first extinction trial revealed a significant main effect of Stimulus, $F(1, 70) = 34.10$, $p < .001$, $\eta^2_p = 0.33$, but no Stimulus × Group interaction, $F(1, 70) = 0.01$, $p = .973$, $\eta^2_p = 0$. These results suggest generalization of the CS+/CS- discrimination to the new context, which was not different between both groups. However, the main effect of Group was significant, $F(1, 70) =$

35.33, $p < .001$, $\eta^2_p = 0.34$, suggesting that the overall US-expectancies are higher in the control group than in the group that received the safety information.

Course of extinction

The middle panel of Figure 1 suggests a decrease in CS+/CS- discrimination from the first to the last extinction trial in both groups with a steeper decline in the control group. A 2 (Stimulus) × 2 (Trial) × 2 (Group) rmANOVA comparing the first and last extinction trial revealed a main effect of Stimulus, $F(1, 70) = 36.94$, $p < .001$, $\eta^2_p = 35$, and a main effect of Trial, $F(1, 70) = 232.74$, $p < .001$, $\eta^2_p = 0.77$. In addition, a significant Stimulus × Trial interaction was found, $F(1, 70) = 12.40$, $p = .001$, $\eta^2_p = 0.15$, but the Stimulus × Trial × Group interaction was not significant, $F(1, 70) = 0.54$, $p = .466$, $\eta^2_p = 0.01$. These results indicate that there was a decrease in CS+/CS- discrimination from the first to the last extinction trial, with no differences between the groups. The Trial × Group interaction was, however, significant, $F(1, 70) = 32.38$, $p < .001$, $\eta^2_p = 0.32$, suggesting that irrespective of CS+/CS- discrimination, a steeper decline from the first to the last extinction trial is observed in the control group compared to the experimental group. Moreover, there was a significant main effect of Group, $F(1, 70) = 12.82$, $p = .001$, $\eta^2_p = 0.16$, indicating lower US-expectancy ratings in the experimental group than in the control group.

Return of fear

To test for group differences in return of fear, a 2 (Stimulus) × 2 (Trial) × 2 (Group) rmANOVA was performed comparing the last extinction trial with the first test trial. This analysis revealed a main effect of Stimulus, $F(1, 70) = 52.69$, $p < .001$, $\eta^2_p = 0.43$, and a main effect of Trial, $F(1, 70) = 125.61$, $p < .001$, $\eta^2_p = 0.64$. The significant Stimulus × Trial interaction, $F(1, 70) = 34.29$, $p < .001$, $\eta^2_p = 0.33$, indicates that there was an increase in CS+/CS- discrimination between the last extinction trial and the first test trial. Moreover, the Stimulus × Trial × Group interaction was significant, $F(1, 70) = 4.03$, $p = .049$, $\eta^2_p = 0.05$, suggesting more return of fear in the control group compared to the experimental group. In addition, there was a significant Trial × Group interaction, $F(1, 70) = 42.01$, $p < .001$, $\eta^2_p = 0.38$. This result indicates that participants who received the safety information, irrespective of CS+/CS- discrimination, showed a smaller increase in US-expectancies from the last extinction to the first test trial compared to participants in the control group.

Course of the test phase

Group differences in the course of the test phase were tested by a 2 (Stimulus) × 3 (Trial) × 2 (Group) rmANOVA including the three test trials. This resulted in a main effect of Stimulus, $F(1, 70) = 64.68$, $p < .001$, $\eta^2_p = 0.48$, and a main effect of Trial, $F(2, 140) = 41.92$, $p < .001$, $\eta^2_p = 0.38$. The significant Stimulus × Trial interaction suggests that there was re-extinction during the test phase, $F(2, 140) = 10.40$, $p < .001$, $\eta^2_p = 0.13$. The Stimulus × Trial × Group interaction was not significant, indicating that both groups did not differ in the course of extinction during the test phase, $F(2, 140) = 0.72$, $p = .449$, $\eta^2_p = 0.01$. However, when looking at the overall US-expectancies and not taking into account CS+/CS- discrimination, the decrease throughout test is steeper in the control group compared to the experimental group, as indicated by a significant Trial × Group interaction, $F(2, 140) = 9.96$, $p < .001$, $\eta^2_p = 0.13$.

Conclusion and Discussion

In clinical practice, patients sometimes receive (objective) information about the probability of the occurrence of their feared outcome as a part of psychoeducation and to lower the threshold to engage in exposure therapy. It has, however, been proposed that giving this type of information might be deleterious to inhibitory learning and to the effectiveness of exposure because it interferes with the possibility to maximally violate expectancies about the occurrence of the aversive outcome during the exposure (Craske et al., 2014). In the present study, we investigated the effect of safety information given between acquisition and extinction training on the return of conditioned fear.

Using an ABA-renewal paradigm, half of the participants (i.e., the experimental group) received between acquisition and extinction the information that the probability of US-occurrence would be extremely small in the remainder of the experiment. The control group did not receive this information. In line with ILT, it was predicted that participants in the experimental group would show higher contextual renewal in the US-expectancy ratings and skin conductance response (SCR) compared to the control group.

The SCR measure did not produce usable results, because during acquisition no differentiation in skin responding between the CS+ and CS- was found, which makes interpretation of the extinction and return of fear data impossible. In the US-expectancy ratings, lower return of fear was observed in the experimental group compared to the control group. The same result was found when taking into account the overall US-expectancy ratings,

irrespective of CS+/CS- discrimination. In addition, whereas it was predicted that the information about US-occurrence would immediately affect US-expectancy ratings, results indicate no group differences in CS+/CS- discrimination on the first extinction trial. However, the information about US-occurrence had the intended effect on the overall US-expectancy ratings (irrespective of CS+/CS- discrimination), with significantly lower ratings on the first extinction trial in the experimental group compared to the control group[2].

Translating these results to clinical practice, we did not find evidence for a deleterious effect of providing information about the (low) probability of the feared outcome on the effectiveness of exposure. Our results even point towards beneficial effects of this type of psychoeducation and suggest that it could attenuate return of fear. Notably, this is opposite to what is predicted by ILT. However, to test more specifically whether the instruction indeed results in less expectancy violation and whether this mediates the outcome of the exposure, a manipulation check might be informative in future research on this topic. For instance, after each extinction trial the degree of surprise that the US did not occur can be assessed (Craske et al., 2014).

There are two accounts of fear learning at the mental level: dual-process models and single-process models (Mertens, Boddez, Sevenster, Engelhard, & De Houwer, 2018). Arguably, ILT and its predictions regarding the effects of safety information before exposure (implicitly) depart from a dual-process account on fear learning. In particular, ILT seems to consider learning via direct experience as opposed and superior to learning via instruction in the sense that their recommendations are focused on preventing that learning via instruction (psychoeducation about US-occurrence) can interfere with the opportunity to subsequently learn via experience in exposure therapy. Such dual-process perspective is at contrast with single-process theories according to which fear learning through verbal instructions and through CS-(no)US pairings are mediated by the same mental process (Mertens et al., 2018). For example, according to propositional learning theory, it does not matter whether information about (the absence of) contingencies is gained by actual experience, instructions or still other pathways (Boddez, De Houwer, & Beckers, 2017; Mitchell, De Houwer, & Lovibond, 2009). More precisely, this theory holds that information from different learning pathways is continuously integrated into one learning process and therefore has no problem in accounting

[2] Notably, results indicate that the safety information also had an effect on the CS- ratings. Whereas, similar to other renewal studies (e.g., Haesen & Vervliet, 2014), an increase in CS- ratings is observed after context change in the control group, this does not seem to be the case in the experimental group.

for our observation that verbally transmitted safety information enhances rather than impedes extinction learning.

Although our findings are at odds with the predictions of ILT, they are in line with previous empirical findings. In a study with a different research question, different instruction, and a different test for return of fear, Sevenster, Beckers, and Kindt (2012) instructed half of their participants after fear acquisition that the CS would no longer be followed by the US, whereas the other participants did not receive these instructions. Both groups then underwent extinction and were tested the next day for return of fear using a reinstatement procedure. Similar to the findings of the current study, the group that received information about the absence of the US showed lower return of fear in the US-expectancy ratings than the control group.

In conclusion, the current study did not find evidence that information about the (low) probability of US-occurrence has deleterious effects on return of fear. These results suggest that in clinical practice providing psychoeducation about the occurrence of the feared outcome does not have negative consequences for the effectiveness of exposure. An important next step is to test this question in a clinical trial.

References

Arch, J. J., Twohig, M. P., Deacon, B. J., Landy, L. N., & Bluett, E. J. (2015). The credibility of exposure therapy: Does theoretical rationale matter? *Behaviour Research and Therapy, 72,* 81-92. doi:10.1016/j.brat.2015.05.008

Bluett, E. J., Landy, L. L., Twohig, M. P., & Arch, J. J. (2016). Does the theoretical perspective of exposure framing matter? Acceptance, fear reduction/cognitive reappraisal, and values-framing of exposure for social anxiety. *Journal of Cognitive Psychotherapy, 30,* 77-93. doi:10.1891/0889-8391.30.2.77

Boddez, Y., Baeyens, F., Hermans, D., & Beckers, T. (2013). Reappraisal of threat value: Loss of blocking in human aversive conditioning. *The Spanish Journal of Psychology, 16,* E84. doi:10.1017/sjp.2013.84

Boddez, Y., Baeyens, F., Hermans, D., Van der Oord, S., & Beckers, T. (2013). Increasing the selectivity of threat through post-training instructions: Identifying one stimulus as a source of danger reduces the threat value of surrounding stimuli. *Journal of Experimental Psychopathology, 4,* 315-324. doi:10.5127/jep.028512

Boddez, Y., De Houwer, J., & Beckers, T. (2017). The inferential reasoning theory of causal learning: Towards a multi-process propositional account. In M. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 1-22). Oxford: Oxford University Press.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114,* 80-99. doi:10.1037/0033-2909.114.1.80

Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear: Tests for the associative value of the context. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 248-265. doi:10.1037/0097-7403.9.3.248

Choy, Y., Fyer, A. J., & Lipsitz, J. D. (2007). Treatment of specific phobia in adults. *Clinical Psychology Review, 27,* 266-286. doi:10.1016/j.cpr.2006.10.002

Clark, D. A., & Beck, A. T. (Eds.). (2010). *Cognitive therapy of anxiety disorders.* New York, NY: The Guilford Press.

Cox, B. J., Endler, N. S., Lee, P. S., & Swinson, R. P. (1992). A meta-analysis of treatments for panic disorder with agoraphobia: Imipramine, alprazolam, and in vivo exposure. *Journal of Behavior Therapy and Experimental Psychiatry, 23,* 175-182. doi:10.1016/0005-7916(92)90034-G

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, N., Chowdhury, N., & Baker, A.

(2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy, 46,* 5-27. doi: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. doi:10.1016/j.brat.2014.04.006

Feske, U., & Chambless, D. L. (1995). Cognitive behavioral versus exposure only treatment for social phobia: A meta-analysis. *Behavior Therapy, 26,* 695-720. doi:10.1016/S0005-7894(05)80040-1

Foa, E. B., Liebowitz, M. R., Kozak, M. J., Davies, S., Campeas, R., Franklin, M. E., …, & Tu, X. (2005). Randomized, placebo-controlled trial of exposure and ritual prevention, clomipramine, and their combination in the treatment of obsessive-compulsive disorder. *American Journal of Psychiatry, 162,* 151-161. doi:10.1176/appi.ajp.162.1.151

Haby, M. M., Donnelly, M., Corry, J., & Vos, T. (2006). Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: A meta-regression of factors that may predict outcome. *Australian and New Zealand Journal of Psychiatry, 40,* 9-19. doi:10.1111/j.1440-1614.2006.01736

Haesen, K., & Vervliet, B. (2015). Beyond extinction: Habituation eliminates conditioned skin conductance across contexts. *International Journal of Psychophysiology, 98,* 529-534. doi:10.1016/j.ijpsycho.2014.11.010

Hembree, E. A., Foa, E. B., Dorfan, N. M., Street, G. P., Kowalski, J., & Tu, X. (2005). Do patients drop out prematurely from exposure therapy for PTSD. *Journal of Traumatic Stress, 16,* 555-562. doi:10.1023/B.JOTS.0000004078.93012.7d

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry, 69,* 621-632.

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: retrospective revaluation induced by experience and instruction. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 97-106. doi:10.1037/0278-7393.29.1.97

Mertens, G., Boddez, Y., Sevenster, D., Engelhard, I. M., & De Houwer, J. (2018). A review on the effects of verbal instructions in human fear conditioning: Empirical findings, theoretical considerations, and future directions. *Biological Psychology*.

Mertens, G., & De Houwer, J. (2016). The impact of a context switch and context instructions

on the return of verbally conditioned fear. *Journal of Behavior Therapy and Experimental Psychiatry, 51*, 10-18. doi:10.1016/j.jbtep.2015.11.001

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*, 183-198.

Norr, A. M., Gibby, B. A., & Schmidt, N. B. (2017). Is computerized psychoeducation sufficient to reduce anxiety sensitivity in an at-risk sample?: A randomized trial. *Journal of Affective Disorders, 212*, 48-55. doi:10.1016/j.jad.2017.01.032

Norton, P. J., & Price, E. C. (2007). A meta-analytic review of adult cognitive-behavioral treatment outcome across the anxiety disorders. *The Journal of Nervous and Mental Disease, 195*, 521-531. doi:10.1097/01.nmd.0000253843.70149.9a

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

Sevenster, D., Beckers, T., & Kindt, M. (2012). Instructed extinction differentially affects the emotional and cognitive expression of associative fear memory. *Psychophysiology, 49*, 1426-1435. doi:10.1111/j.1469-8986.2012.01450.x

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*, 36-45. doi:10.1027/1618-3169/a000005

Stanley, M. A., & Turner, S. M. (1995). Current status of pharmacological and behavioral treatment of obsessive-compulsive disorder. *Behavior Therapy, 26*, 163-186. doi:10.1016/S0005-7894(05)80089-9

Van Etten, M. L., & Taylor, S. (1998). Comparative efficacy of treatment for post-traumatic stress disorder: A meta-analysis. *Clinical Psychology & Psychotherapy, 5*, 126-144. doi:10.1002/(SICI)1099-0879(199809)5:3<126::AID-CPP153>3.0.CO;2-H

Vander Haegen, M., & Etienne, A.-M. (2016). Cognitive processes across anxiety disorders related to intolerance of uncertainty: Clinical review. *Cogent Psychology, 3*, 1215773. doi:10.1080/23311908.2016.2115773

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour Research and Therapy, 43*, 357-371. doi:10.1016/j.brat.2004.02.005

Whittal, M. L., Thordarson, D. S., & McLean, P. D. (2005). Treatment of obsessive-

compulsive disorder: Cognitive behavior therapy vs. exposure and response prevention. *Behaviour Research and Therapy, 43*, 1559-1576. doi:10.1016/j.brat.2004.11.012

Chapter 6

The validity of laboratory-based treatment research: Bridging the gap between fear extinction and exposure treatment

Based on:

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

Abstract

A major objective of experimental psychopathology research is to improve clinical practice via the experimental study of treatment mechanisms. The success of this endeavor depends on the external validity of the procedures used to model the treatment component in the laboratory. We propose a general framework and a set of specific criteria that will allow evaluating whether a certain laboratory procedure is a valid model for a certain clinical treatment. We illustrate this framework by evaluating the validity of extinction as a laboratory model for clinical exposure therapy. Although we acknowledge the merits of the extinction model, we argue that its validity might not be as firmly established as the research community assumes. We also use extinction as an example to demonstrate how considerations of the proposed criteria can stimulate further improvements to existing models of treatment. We conclude that the systematic assessment of external validity of treatment models is an important step towards bridging the gap between science and practice in the field of experimental psychopathology.

Keywords: Fear extinction; Exposure therapy; Return of fear; Validity; Experimental psychopathology; Translational research

The validity of laboratory-based treatment research: Bridging the gap between fear extinction and exposure treatment

Experimental psychopathologists study the causal factors of pathological behavior under highly controlled conditions. According to Kimmel (1971), experimental psychopathology (EP) can be approached as both "the experimental study of pathological behavior" and "the study of experimental pathological behavior" (p. 7, see also Forsyth & Zvolensky, 2002; Zvolensky, Forsyth, & Johnson, 2013). The former approach concerns the experimental study of (factors that influence) pre-existing pathological behavior in clinical or subclinical subjects. In the latter approach of EP, 'pathological behavior' is experimentally induced in healthy (animal or human) subjects. A prerequisite for research in healthy subjects is a laboratory model of the pathological behavior: a set of behavioral, pharmacological, genetic or surgical manipulations that result in behavior that is similar to the pathological behavior. Pavlov (1927), to give an early behavioral example, produced behavior similar to neurosis by presenting his dogs with ambiguous stimuli. In a first phase of a relatively easy discrimination task, a circle but not an ellipse was presented together with food. Subsequent presentation with a stimulus somewhere in between a circle and an ellipse resulted in symptoms characteristic of neurosis. An example of a genetic manipulation is the cannabinoid receptor gene knockout mouse that exhibits behavioral changes that are similar to symptoms of schizophrenia (Fritzsche, 2001). Once the prerequisite of having a laboratory model of the pathological behavior is met, a plethora of research questions can be investigated (e.g., about individual differences or about the environmental factors that exacerbate such behavior; Vervliet & Raes, 2013), and hence a better understanding of this behavior can be attained.

However, the ultimate aim of experimental psychopathologists is not to merely understand, but also to reduce pathological behavior. Despite a great amount of EP research, there are still opportunities for the further enhancement of clinical treatments. Only about half of the patients experience a full remission or respond to psychological treatment in a clinically meaningful way (Holmes, Craske, & Graybiel, 2014). Moreover, an important subgroup of patients fails to maintain the effects of treatment in the long term and experiences relapse (e.g., Lipsitz, Mannuzza, Klein, Ross, & Fyer, 1999; Steinert, Hofmann, Kruse, & Leichsenring, 2014).

A straightforward factor that might add to the continued development of clinical treatment is more interaction between scientists and clinicians (e.g., Barlow, 1981; Berke, Rozell, Hogan, Norcross, & Karpiak, 2011). In line with this, evidence-based strategies to disseminate and

implement evidence-based interventions have recently started to develop (McHugh & Barlow, 2010). In addition to enhancing communication, investing in the external validity of treatment models provides an opportunity to further improve clinical treatment. In laboratory research, complex psychological treatments are reduced to the putative core mechanisms (e.g., van den Hout, 1999). Such reduction contributes to the internal validity of the model: by providing control over confounding variables, reliable causal inferences can be made (van den Hout, Engelhard, & McNally, in press). This is indeed considered one of the major strengths of EP research. Also, from a pragmatic point of view, it is more cost-effective and less time consuming to first test hypotheses in healthy volunteers using a basic treatment model before testing them in clinical trials. The question is, however, whether findings obtained with these simplified treatment models are still informative for clinical practice. In the present paper, we propose a general framework to answer this question. In particular, we discuss three criteria that have long been outlined in pharmacological research and have recently been used to evaluate the external validity of experimental models for psychopathology: face validity, construct validity and predictive validity (e.g., Abramson & Seligman, 1977; Boddez et al., 2013; Luyten, Vansteenwegen, Van Kuyck, Gabriels, & Nuttin, 2011; Vervliet & Raes, 2013). To the best of our knowledge, this is the first time that these validity criteria are applied to a psychological treatment model. We illustrate this framework by evaluating the validity of extinction as a treatment model for clinical exposure therapy.

Fear extinction is seen as one of the most successful treatment models in the history of EP (Vervliet, Craske, & Hermans, 2013). Its laboratory procedure entails unreinforced presentations of the conditioned stimulus (CS; e.g., geometrical shape), resulting in a decrease in the fear responses that were previously established by pairing the CS with an aversive (unconditioned) stimulus (US; e.g., electrical shock). This procedure is used to model clinical exposure therapy (e.g., Craske, Hermans, & Vansteenwegen, 2006). In exposure-based treatments, the anxious client is repeatedly and systematically confronted with the fear-provoking situation (e.g., McNally, 2007). Despite being an efficacious treatment for a range of anxiety disorders, relapse is not uncommon after exposure-based treatments (e.g., Simpson et al., 2004). Limited generalization of extinction is generally considered to be the preeminent laboratory model for relapse following exposure therapy (e.g., Bouton, 2002). But how can we know whether continued research into fear extinction will teach us more about exposure treatment and ways to improve it? This question is fundamental to the issue of external validity and speaks directly to the challenge of bridging the gap between science and treatment.

Below, we discuss each of the three validity criteria (face validity, construct validity and predictive validity) in separate sections. We start each section with a definition of the criterion as applied to treatment models. Subsequently, we evaluate the extinction model using this criterion. We end each section by using extinction as an example to demonstrate how the validity approach can guide future developments in laboratory-based treatment research.

Face validity

Definition

In the present context, face validity refers to the surface similarity between the treatment model and the treatment itself. Surface similarity (face validity) is generally seen as a good starting point for the development of experimental models, but it is deemed as not very informative for the external validity of a model (e.g., Vervliet & Raes, 2013). That is because mere similarities in procedure or result, however compelling, do not imply that similar mechanisms are involved (i.e., construct validity) or that treatment enhancing strategies that prove to be successful in the laboratory will also be successful in clinical practice (i.e., predictive validity). Nevertheless, surface similarity with clinical treatment does remain important, because it can serve as a continuing source of inspiration for creating new laboratory models or updating existing ones.

Extinction and return of fear

We now turn to the assessment of the extinction model using this criterion. Many researchers do refer to surface similarity when justifying their choice for fear extinction as a model of exposure treatment, as evidenced by the introduction sections of many published studies on extinction (e.g., Culver, Vervliet, & Craske, 2015; Kindt & Soeter, 2013; Leer & Engelhard, 2015). In both extinction training and exposure-based treatment, the repeated confrontation with a fear-evoking situation or stimulus results in a decrease in outcome variables that are indicative of fear and anxiety (e.g., US-expectancy, subjective units of distress ratings). The same holds for laboratory models of relapse. Return of fear is a well-documented phenomenon after fear extinction in the laboratory (Vervliet et al., 2013). Two paradigms frequently used for this purpose are renewal and reinstatement (Vervliet, Baeyens, Van den Bergh, & Hermans, 2013). In renewal, a context switch between the extinction phase and the test phase causes a return of fear responses similar to a clinical relapse after successful treatment when the feared object or situation is encountered outside the therapy context (Effting & Kindt, 2007).

Reinstatement refers to the return of fear after unsignaled US-presentations between extinction and test, and can be seen as the equivalent of relapse after unsignaled panic attacks or if the previously feared stimulus is encountered after a stressful event or in a distressing situation (Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2007; Haaker, Golkar, Hermans, & Lonsdorf, 2014). In conclusion, at face value fear extinction seems to be a sufficiently good treatment model of exposure therapy.

Future research

Researchers can continue to invest in increasing the surface similarity between the extinction procedure and exposure treatment. For example, it has been argued that basic stimulus sets such as geometrical shapes lack the complexity of real-world experiences (e.g., Barry, Griffith, De Rossi, & Hermans, 2014). Some researchers therefore turn to the use of 3-D virtual reality technology that allows administering extinction training under conditions that are closer to real-life situations (e.g., Dunsmoor, Ahs, Zielinski, & LaBar, 2014). Using more complex, multi-sensory stimuli (e.g., auditory, tactile, olfactory, visual) can be a conceivable step in increasing the procedural overlap between extinction and exposure therapy (for a similar argument, see Waters, LeBeau, & Craske, in press). In addition, extensions of the extinction model aimed at improving its face validity can target similarities in outcome measures. Behavioral avoidance is an important source of impairment in daily functioning in pathological anxiety (Barlow, 2002) and is frequently used as an outcome measure in clinical exposure studies, by using a behavioral approach task (e.g., Niles, Craske, Lieberman, & Hur, 2015). The external validity of extinction research might therefore benefit from including behavioral avoidance as an outcome measure in addition to expectancy or fear ratings and psychophysiological indices of fear (e.g., van Meurs, Wiggert, Wicker, & Lissek, 2014; Vervliet & Indekeu, 2015).

However, as mentioned before, enhancing the surface similarity of the extinction model does not by itself imply enhanced external validity with regard to underlying mechanisms or enhanced predictive value. Future research should therefore verify whether extensions of the extinction model aimed at increasing face validity do add to the transfer of successful interventions investigated in the laboratory to clinical exposure therapy. The validity criteria that are typically considered more decisive — construct validity and predictive validity — are discussed in the next sections.

Construct validity

Definition

The criterion of construct validity is met if the psychological mechanisms that drive behavioral change in the treatment model are the same as those that drive behavioral change in clinical treatment. Making this comparison requires a profound theoretical understanding of the mechanisms at play in the treatment model and in clinical treatment. Without an elaborate theoretical ground, the model might resemble the clinical treatment in terms of surface similarity (i.e., face validity), but further developments to optimize the clinical treatment based on the model will be hampered. That will, for example, be the case if an intervention that shows promise in the laboratory model exerts its effect through targeting a cognitive process that is at play in the model, but not in the clinical treatment. In such case the promise of the intervention will not be fulfilled in clinical treatment. In addition, construct validity is essential for the generation of theory-based optimization strategies.

Extinction and return of fear

Extinction. The currently dominant theory attributes extinction effects to an inhibitory learning mechanism (e.g., Bouton, 2002). The central tenet of this theory is that, rather than a destruction of the original excitatory CS-US association, an additional inhibitory CS-noUS association is learned throughout extinction training. This inhibitory association cancels out the original excitatory association, resulting in low to no conditioned responding when both these associations are activated. The theory additionally assumes that retrieval of the inhibitory association is modulated by the context, which allows it to explain the limited generalization of extinction performance. The inhibitory learning approach has proven its merits in accounting for extinction phenomena at the behavioral level (e.g., Bouton, 1993) and did find support in neurobiological research as well. As a result of extinction training, it is observed that prefrontal areas show an increase in activity, whereas areas activated during fear acquisition (e.g., the amygdala) show a decrease in activity, often interpreted as the former areas inhibiting the latter (e.g., Milad et al., 2007; Milad et al., 2005; Milad & Quirk, 2012; Phelps, Delgado, Nearing, & LeDoux, 2004).

High construct validity of the extinction treatment model would imply that a similar inhibitory mechanism drives the effects of exposure therapy. Speaking to its success, the

inhibitory learning approach has provided useful recommendations for conducting clinical exposure therapy (see Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; Pittig, van den Berg, & Vervliet, 2016). Rather than pursuing fear reduction during treatment, which is a central component in habituation-based theories of exposure therapy such as the Emotional Processing Theory (e.g., Foa & Kozak, 1986), the inhibitory learning approach focuses on enhancing inhibitory learning and facilitating the retrieval of the inhibitory association following successful treatment (Craske et al., 2008; Craske, Liao, Brown, & Vervliet, 2012). Theoretically speaking, acquisition of an inhibitory association can be strengthened by, for example, maximizing expectancy violation, deepened extinction (i.e., multiple CS's are presented together during extinction after being extinguished separately; Culver, et al., 2015), and occasional reinforcement during extinction (Craske et al., 2014). Retrieval of an inhibitory association can be facilitated, amongst others, by the use of retrieval cues (e.g., Vansteenwegen et al., 2006) and by inducing variability in timing, stimuli and contexts (e.g., Bandarian-Balooch, Neumann, & Boschen, 2012). Although there is a need for further evidence, some of these strategies have proven successful in optimizing exposure therapy, hence providing support for a role of the inhibitory learning mechanism in exposure therapy (e.g., Mystkowski, Craske, Echiverri, & Labus, 2006; Rowe & Craske, 1998$_a$; Tsao & Craske, 2000; Vansteenwegen et al., 2007). Furthermore, neuroimaging research suggests that the neural mechanisms underlying exposure therapy are similar to those supporting inhibitory learning as a crucial mechanism underlying extinction. In particular, a normalization of amygdala responses is observed after exposure therapy (Goossens, Sunaert, Peeters, Griez, & Schruers, 2007), which is modulated by increased prefrontal activity (e.g., Felmingham et al., 2007; Hauner, Mineka, Voss, & Paller, 2012). Felminghan, Dobson-Stone, Schofield, Quirk, and Bryant (2012) found that BDNF genotype, associated with reduced inhibition of the amygdala by prefrontal areas, predicts poor exposure treatment response. These results are consistent with findings from extinction research (Soliman et al., 2010).

At the same time, exposure therapy most likely encompasses more than the acquisition of an inhibitory association (e.g., Carey, 2011; Hofmann, 2008; Lovibond, 2004). Among other things, real-life exposure treatment might also recruit from other processes studied in the associative learning tradition, including US-habituation, US-devaluation and counterconditioning (e.g., Jaycox, Foa, & Morral, 1998; Tyron, 2005). Although exposure to the US is unethical in some cases, two notable exceptions are the treatment of social anxiety and panic disorder. In the treatment of social anxiety, the client can be exposed to social

rejection (Craske et al., 2014). In the treatment of panic disorder, the feared physical sensations (e.g., hyperventilation, increased heart rate, dizziness) are deliberately induced during interoceptive exposure, and this occasionally results in a full-blown panic attack (Craske & Barlow, 2014). By experiencing panic attacks during treatment and in the presence of the therapist, clients can come to evaluate a panic attack as less dangerous and aversive (i.e., US-revaluation; Rescorla, 1973). Reinforcing feedback and compliments by the therapist and surrounding others might also result in counterconditioning: the feared stimulus is presented together with a positive event (i.e., compliments) that might change the valence of the feared stimulus (De Houwer & Hughes, in press).

In addition, whereas extinction is typically investigated in a Pavlovian procedure in which the participant is passively exposed to the CS, operant processes most likely play an important role in exposure-based treatments (Bouton & Todd, 2014). During clinical exposure, clients not only learn to refrain from avoidance behaviors, but are also encouraged to actively approach the feared situation. This approach behavior might add to the effects of exposure in several ways. First, merely approaching an object leads to more positive attitudes towards that object (Jones, Vilensky, Vasey, & Fazio, 2013). Second, clients learn to acquire control over their behavior despite being anxious and experience that they are able to cope with the feared situation, leading to an increased sense of self-efficacy (e.g., Tryon, 2005). This operant component is no part of the extinction treatment model, although it is considered to be one of the primary therapeutic targets by many therapists in clinical practice (Hayes, Strosahl, & Wilson, 2003, but see van Uijen, van den Hout, and Engelhard, 2015).

It can be concluded that the construct validity of extinction as a model of exposure therapy is not unequivocally supported and needs to be elucidated more clearly. While acknowledging the considerable success of the inhibitory learning account, we also argued that there are mechanisms that are at work in exposure treatment but not in the extinction model, including US-habituation, counterconditioning and approach behavior (McConnell & Miller, 2014).

Return of fear. We now extend our analysis from extinction as a model of exposure therapy to return of fear as a model of relapse. Based on laboratory research on return of fear, the research community aims to draw conclusions about relapse and its prevention in clinical practice. Accordingly, it is important that the return of fear observed after extinction training is mediated by similar mechanisms as relapse after exposure treatment. So, the question becomes: are similar mechanisms responsible for the return of extinguished fear responses and for clinical

relapse? Processes at work in exposure therapy, but not in the extinction treatment model, might exert influences on the generalization of the behavioral change. The small amount of available evidence suggests that some of these processes —more specifically, US-habituation and counterconditioning— are better retained over time and are less vulnerable to generalization decrement than extinction training (Dibbets, Poort, & Arntz, 2012; Haesen & Vervliet, 2015; Kerkhof, Vansteenwegen, Baeyens, & Hermans, 2010; but see Bouton, 2002; Brooks, Hale, Nelson, & Bouton, 1995). Haesen and Vervliet (2015), for instance, found that participants who were exposed to the US (i.e., a procedure to obtain US-habituation) showed less renewal in skin-conductance response compared to participants who were exposed to the CS (i.e., standard extinction training). Isolating the inhibitory learning component of exposure therapy in laboratory research might therefore result in ignoring the interplay between such processes and in an incomplete understanding of how the effects of exposure therapy generalize.

Another potential confound in translating laboratory findings on return of fear to clinical practice is that, arguably, laboratory phenomena such as renewal and reinstatement are adequate models for temporarily re-experiencing fear in real life, but not necessarily for clinical relapse. In the laboratory, return of fear is typically defined as the increase in US-expectancy and physiological fear responses between the end of extinction training and the first trial of an unreinforced test phase (e.g., Barry, Griffith, Vervliet, & Hermans, 2016). However, return of fear observed in the laboratory is typically not persistent: immediately after the first test trial, a steep decline in the 'returned' fear responses is observed (e.g., Barry et al., in press; Vervliet, Kindt, Vansteenwegen, & Hermans, 2010). Therefore, phenomena such as renewal and reinstatement might correspond to a short-term recurrence of fear or a lapse rather than a full-blown relapse (i.e., a long-term recurrence of the full symptomatology). Although in clinical practice a lapse can be a precursor for a full relapse, not all clients who experience a provisional recurrence of fear after treatment actually relapse (Vervliet et al., 2013). A critical mediator in the transition between a lapse and an actual relapse could be re-engaging in excessive and rigid avoidance behaviors in response to a recurrence of fear. Similar to its role in the etiology and maintenance of anxiety disorders, it seems reasonable to assume that rigid avoidance in response to a short-term recurrence of fear after treatment takes away the opportunity to experience that the expected or feared outcomes do not occur and that one is able to cope with the fear-eliciting situation (Forsyth, Eifert, & Barrios, 2006; Lommen, Engelhard, Sijbrandij, van den Hout, & Hermans, 2013; Mineka & Zinbarg, 2006). Notably, models for relapse typically used in the laboratory are restricted to a Pavlovian procedure and do not include an

opportunity to perform avoidance behavior when experiencing return of fear at the first test trial. Because participants cannot escape from or avoid the CS in these procedures, they will experience that the US does not occur and immediate (re-)extinction will take place after the first test trial.

Admittedly, because of ethical considerations laboratory research will always be limited in its ability to model clinical relapse. However, below we illustrate how feasible adjustments to the extinction model could result in a laboratory model of relapse that might better reflect clinical relapse.

Future research

A number of suggestions to optimize the extinction model of exposure therapy can be derived from the evaluation of its construct validity. We discussed that additional processes, disregarded in current extinction research, might be at work in exposure therapy and that some of these processes are less subject to generalization decrement. An obvious research strategy would be to form hypotheses about the various mechanisms at work in exposure therapy and to investigate whether they show effects on extinction learning and its generalization. Let us illustrate this with the example of operant processes. As said, patients are encouraged to actively approach the feared situation during exposure, which might lead to more positive attitudes towards the feared object (Jones et al., 2013) and to an increased sense of self-efficacy (e.g., Tryon, 2005). In future research, one could add an approach component by having participants pull the CS towards them with a joystick during the extinction phase (with the CS enlarging in response to the joystick movement; Krypotos, Arnaudova, Effting, Kindt, & Beckers, 2015) and assess whether that changes the course of extinction and its generalization. We also argued that return of fear observed in laboratory procedures is typically not persistent due to immediate (re-)extinction. Including the possibility to avoid the CS or escape from the CS when testing the generalization of extinction might better reflect what happens in clinical practice and could result in more persistent return of fear.

The decision to update a standard laboratory model should, however, depend on whether including such extra components adds to the successful transfer of laboratory findings to clinical exposure therapy (i.e., the predictive validity). For example, an extinction model that includes these operant components should only replace the standard extinction model if it allows drawing stronger conclusions of clinical significance. We further elaborate on this type of validity in the next section.

Predictive validity

Definition

Whereas construct validity concerns the question whether or not the same theoretical mechanisms are at work in the model and in the clinical treatment, predictive validity is a criterion that relates to the question whether performance in the model predicts performance in the clinical treatment. Thus claims about latent psychological mechanisms are at the heart of construct validity, whereas predictive validity in principle remains silent with respect to these mechanisms.

We distinguish two aspects in evaluating the predictive validity of psychological treatment models. A first aspect refers to testing whether individual differences in performance in the model predict the course and outcome of the clinical treatment. According to this aspect, a model has predictive validity if individuals who perform better in the model also benefit more from the clinical treatment. A second aspect of evaluating the predictive validity concerns testing whether interventions and factors that influence behavior in the model also influence the clinical treatment and vice versa. This aspect is directly attached to the translational value of the model.

Extinction and return of fear

Testing the first aspect of predictive validity requires investigating whether individual extinction performance is correlated with performance in exposure therapy. An important step would be to assess whether individuals who show steeper extinction curves in the laboratory also show greater decreases in indices of fear and anxiety during exposure treatment (i.e., correlation between the course of extinction and exposure). Similarly, individuals who show incomplete extinction would be expected to experience more residual anxiety at the end of treatment (i.e., correlation between the immediate outcome of extinction and exposure). Furthermore, the predictive validity of laboratory models for relapse can be tested by examining whether the amount of return of fear observed in the model is correlated with relapse after exposure therapy. Surprisingly, these tests have not been carried out in a systematic way.

The second aspect of predictive validity involves testing whether interventions of interest have similar effects on performance in the extinction model and exposure therapy. This aspect of predictive validity can be demonstrated in two directions: from the clinical treatment to the

model and from the model to the clinical treatment. The former involves showing that factors that are known to influence clinical exposure treatment influence extinction training in a similar way. The latter involves the successful translation of interventions first tested in the extinction model to exposure therapy. Below, we discuss research on self-efficacy and D-cycloserine to illustrate both directions of this type of predictive validity.

Self-efficacy or the belief that one can effectively cope with fear-eliciting situations, is an important determining factor for both immediate and long-term effects of exposure therapy (e.g., Goldin et al., 2005). In particular, increases in self-efficacy have been found to mediate the effects of exposure-based treatment. Similar effects of self-efficacy on extinction would provide an argument for its predictive validity. A recent study indeed showed that adding verbal persuasion aimed at increasing self-efficacy in a human fear conditioning paradigm facilitated extinction of fear (Zlomuzica, Preusser, Schneider, & Margraf, 2015).

The second example illustrates predictive validity in the other direction, from the extinction model to exposure therapy. Administering D-cycloserine (DCS), a partial agonist at the glycine-binding site of the N-methyl-D-aspartate receptor in the amygdala, has been shown to facilitate fear extinction and strengthen extinction memory in rodents (e.g., Vervliet, 2008; Walker & Davis, 2002). If extinction is a valid model for exposure therapy, adding DCS should improve the outcome of clinical exposure therapy in a similar way. This was confirmed in several meta-analyses (Bontempo, Panza, & Bloch, 2012; Norberg, Krystal, & Tolin, 2008; Rodrigues et al., 2014; but see Hofmann et al., 2013; Ori et al., 2015).

Future research

There is plenty of room for future research inspired by the evaluation of the predictive validity of extinction as a model of exposure therapy. The most important challenge for future research remains to test whether individual differences in the course and outcome of extinction training can predict the course and outcome of clinical exposure treatment. These tests are a relatively straightforward way to assess the predictive validity of extinction.

With regard to the second aspect of predictive validity, a substantial subset of extinction-enhancing strategies have not been tested systematically in exposure therapy (e.g., Craske et al., 2014; Pittig, et al., 2016). Future translational research is warranted to test the effects of strategies such as deepened extinction (Culver et al., 2015) and occasional reinforcement (Culver, 2013) in clinical exposure studies. In the other direction, it can be examined whether

variables and strategies that have proven successfully in optimizing exposure therapy, such as affect labeling (Kircanski, Lieberman, & Craske, 2012; Niles et al., 2015) and expanding-spaced exposure schedules (Rowe & Craske, 1998$_b$; Tsao & Craske, 2000) can attenuate return of fear in human fear extinction as well.

The second aspect of predictive validity involves bridging the gap between laboratory techniques and clinical practice. It is, however, important to note that, with regard to our present purposes, the goal of these research studies would primarily be to assess the external validity of the extinction model. That is, we here argue for these translational studies because they are a means to evaluate whether extinction is a good model for exposure therapy, rather than because they involve translating a specific and promising laboratory technique to clinical practice.

Conclusion and Discussion

The issue of external validity has received ample interest in pharmacological research and some interest in EP research concerned with the modeling of psychopathology (e.g., Abramson & Seligman, 1977; Vervliet & Raes, 2014). The external validity of models of psychological treatment, although no less important, has received less attention. Threatened external validity of treatment models, amongst other factors, can explain why promising findings of experimental research do not generalize to clinical treatment. This has led some clinically-oriented researchers to argue in favor of testing interventions directly in clinical research and skip experimental approaches (Sloan, 2014). However, EP treatment research can make a unique contribution to the advancement of (transdiagnostic) treatment next to clinical studies by offering a profound understanding of treatment mechanisms. To maximize the translational value of EP research, it is, however, important that treatment models are subjected to systematic evaluation with regard to their external validity. In the present paper, we propose a framework for such evaluation. In particular, we demonstrate how the validity criteria previously outlined in pharmacological research (i.e., face validity, construct validity and predictive validity) can be applied to models of psychological treatment. Using extinction as a model for exposure-based therapy, we illustrate how this framework is not only suitable to evaluate treatment models, but can also stimulate future research to optimize models for psychological treatment.

Face validity, or the surface similarity between the treatment model and the clinical treatment, can be considered a good starting point and a source of inspiration for the further development of a model. However, it is not a sufficient ground for the external validity of a

treatment model. We concluded that extinction has sufficient face validity and that even higher face validity could be achieved by including behavioral avoidance as an outcome measure and by making use of more complex, multi-sensory stimuli.

The construct validity of a treatment model depends on whether the mechanisms that drive behavioral change in the model correspond to those underlying the clinical treatment. An elaborate theoretical framework is indispensable for this endeavor. We acknowledge that Bouton's inhibitory learning theory (1993) does provide a data-supported, parsimonious, and simple foundation for understanding extinction and most recovery effects. However, there are reasons to believe that it is not sufficient to tell the whole story when applied to clinical exposure and relapse (McConnell & Miller, 2014). Several additional mechanisms might account for the effects of exposure therapy, including US-habituation, US-devaluation, counterconditioning, and operant learning. Future research could benefit from taking these additional mechanisms into account. The research suggestions that we provided demonstrate that extinction research permits such extension.

The predictive validity of a laboratory model is directly related to its translational nature and therefore of great pragmatic value. A first aspect of the predictive validity is to test whether the course and outcome of the clinical treatment can be predicted by individual differences in performance in the model. An important target for future research in the field of extinction is to carry out these tests in a systematic way. A second aspect of the predictive validity is to investigate whether factors and interventions have a similar effect in the model and the clinical treatment. As said, a substantial subset of manipulations has not been systematically examined with regard to whether they have a similar impact on extinction and exposure therapy.

It is important to note that the three criteria of external validity are closely related. The procedure to which a clinical treatment is reduced in the laboratory is developed on the basis of theoretical assumptions about the core mechanisms at work in treatment. This concerns the construct validity of the treatment model. This reduction is an inherent part of the experimental method used in EP research and enables EP researchers to gain control over confounding variables, thereby contributing to internal validity. At the same time, strong reduction of a clinical treatment might detract from the face validity of a treatment model, whereas investing in face validity might indirectly add to construct and predictive validity. Although there is definitely not a one-to-one relation, it can be assumed that the greater the overlap at face value between the treatment model and the clinical treatment, the greater the chance that the model includes factors or processes that are crucial in the clinical treatment. However, this generally

also results in a more complex laboratory model. The critical test for determining to what extent a clinical treatment can be simplified in the laboratory is the predictive validity of the model (van den Hout, et al., in press). If a highly simplified, basic treatment model has satisfactory predictive value, this basic model has to be preferred above a more complex model with equal predictive validity (cf. Ockham's razor). If, however, excessive reduction significantly dilutes the predictive validity of the treatment model, the more complex model is recommended.

Predictive validity cannot only be used to determine the optimal balance between face validity and construct validity, but is also more directly related to construct validity. In particular, it can be argued that a model that includes the underlying processes of the clinical treatment will also have a greater predictive value. This interaction can again be illustrated with inhibitory learning theory. If inhibitory learning is a crucial process underlying both extinction and exposure therapy, it is more likely that strategies and interventions for enhancing inhibitory learning and its retrieval exert a similar effect on extinction and exposure.

In conclusion, we were inspired by the observation that the progress of psychological treatment is lagging behind theoretical advances and promising findings in the laboratory. We hope that assessment of external validity through the proposed criteria can stimulate the further optimization of treatment models and help to bridge the gap between laboratory-based treatment research and clinical practice.

References

Abramson, L., & Seligman, M. (1977). Modeling psychopathology in the laboratory: history and rationale. In J. Maser & M. Seligman (Eds.), *Psychopathology: Experimental Models* (pp. 1-26). San Francisco: W. H. Freeman and Co.

Bandarian-Balooch, S., Neumann, D., & Boschen, M. J. (2012). Extinction treatment in multiple contexts attenuates ABC renewal in humans. *Behaviour Research and Therapy, 50,* 604-609. doi:10.1016/j.brat.2012.06.003

Barlow, D. H. (1981). On the relation of clinical research to clinical practice: Current issues, new directions. *Journal of Consulting and Clinical Psychology, 49,* 147-155.

Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic*. New York, NY: Guilford Press.

Barry, T. J., Griffith, J. W., De Rossi, S., & Hermans, D. (2014). Meet the Fribbles: Novel stimuli for behavioural research. *Frontiers in Psychology, 5,* 1-8. doi:10.3389/fpsyg.2014.00103

Barry, T, J., Griffith, J. W., Vervliet, B., Hermans, D. (2016). The role of stimulus specificity and attention in the generalization of extinction. *Journal of Experimental Psychopathology, 7,* 143-152. doi:10.5127/jep.048615

Berke, D. M., Rozell, C. A., Hogan, T. P., Norcross, J. C., & Karpiak, C. P (2011). What clinical psychologists know about evidence-based practice: Familiarity with online resources and research methods. *Journal of Clinical Psychology, 67,* 329-339.

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,* 201-206. doi:10.1016/j.jbtep.2012.08.003

Bontempo, A., Panza, K. E., & Bloch, M. H. (2012). D-cycloserine augmentation of behavioral therapy for the treatment of anxiety disorders: A meta-analysis. *The Journal of Clinical Psychiatry, 73,* 533-537. doi:10.4088/JCP.11r07356

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114,* 80-99. doi:10.1037/0033-2909.114.1.80

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioural extinction. *Biological Psychiatry, 52,* 976-986.

Bouton, M. E., & Todd, T. P. (2014). A fundamental role for context in instrumental learning

and extinction. *Behavioural Processes, 104*, 13-19. doi:10.1016/j.beproc.2014.02.012

Brooks, D. C., Hale, B., Nelson, J. B., & Bouton, M. E. (1995). Reinstatement after counterconditioning. *Animal Learning and Behavior, 23*, 383-390. doi:10.3758/BF03198938

Carey, T. A. (2011). Exposure and reorganization: The what and how of effective psychotherapy. *Clinical Psychology Review, 31*, 236-248. doi:10.1016/j.cpr.2010.04.004

Craske, M. G., & Barlow, D. H. (2014). Panic disorder and agoraphobia. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (pp. 1-62). New York, NY: Guilford Press.

Craske, M. G., Hermans, D., & Vansteenwegen, D. (2006). *Fear and learning: From basic processes to clinical implications*. Washington, D. C.: American Psychological Association.

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*, 5-27. doi:10.1016/j.brat.2007.10.003

Craske, M. G., Liao, B., Brown, L., & Vervliet, B. (2012). Role of inhibition in exposure therapy. *Journal of Experimental Psychopathology, 3*, 322-345. doi:10.5127/jep.026511

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. doi:10.1016/j.brat.2014.04.006

Culver, N., C. (2013). *Extinction-based processes for enhancing the effectiveness of exposure therapy*. University of California, Los Angeles.

Culver, N. C., Stoyanova, M. S., & Craske, M. G. (2011). Clinical relevance of retrieval cues for attenuating context renewal of fear. *Journal of Anxiety Disorders, 25*, 284-292. doi:10.1016/j.janxdis.2010.10.002

Culver, N. C., Vervliet, B., & Craske, M. G. (2015). Compound extinction: Using the Rescorla-Wagner model to maximize exposure therapy effects for anxiety disorders. *Clinical Psychological Science, 3*, 335-348. doi:10.1177/2167702614542103

De Houwer, J., & Hughes, S. (in press). Evaluative conditioning as a symbol phenomenon: On the relation between evaluative conditioning, evaluative conditioning with instructions, and persuasion. *Social Cognition.*

Dibbets, P., & Maes, J. H. R. (2011). The effect of an extinction cue on ABA-renewal: Does

valence matter? *Learning and Motivation, 42*, 133-144. doi:10.1016/j.lmot.2010.12.003

Dibbets, P., Poort, H., & Arntz, A. (2012). Adding imagery rescripting during extinction leads to less ABA renewal. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 614-624. doi:10.1016/j.jbtep.2011.08.006

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2007). Reinstatement of conditioned responses in human differential fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 237-251. doi:10.1016/j.btep.2006.04.001

Dunsmoor, J. E., Ahs, F., Zielinski, D. J., & LaBar, K. S. (2014). Extinction in multiple virtual reality contexts diminishes fear reinstatement in humans. *Neurobiology of Learning and Memory, 113*, 157-164. doi:10.1016/j.nlm.2014.02.010

Effting, M, & Kindt, M. (2007). Contextual control of human fear associations in a renewal paradigm. *Behaviour Research and Therapy, 45*, 2002-2018. doi:10.1016/j.brat.2007.02.011

Felmingham, K. L., Dobson-Stone, C., Schofield, P. R., Quirk, G. J., & Bryant, R. A. (2012). The brain-derived neurotrophic factor Val66Met polymorphism predicts response to exposure therapy in posttraumatic stress disorder. *Biological Psychiatry, 73*, 1059-1063. doi:10.1016/j.biopsych.2012.10.033

Felmingham, K. L., Kemp, A. H., Williams, L.M., Das, P., Hughes, G., Peduto, A., & Bryant, R. A. (2007). Anterior cingulate and amygdala changes after cognitive behavioural therapy in posttraumatic stress disorder. *Psychological Science, 18*, 127-129. doi:10.1111/j.1467-9280.2007.01860.x

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin, 99*, 20-35.

Forsyth, J. P., Eifert, G. H., & Barrios, V. (2006). Fear conditioning in an emotion regulation context: A fresh perspective on the origins of anxiety disorders. In: M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.), *From basic processes to clinical implications.* Washington, D. C.: American Psychological Association.

Forsyth, J. P., & Zvolensky, M. J. (2002). Experimental Psychopathology, clinical science and practice: An irrelevant or indispensable alliance? *Applied & Preventive Psychology, 10*, 243-264.

Fritzsche, M. (2001). Are cannabinoid receptor knockout mice animal models for schizophrenia? *Medical Hypotheses, 56*, 638-643. doi:10.1054/mehy.2000.1261

Goldin, P. R., Ziv, M., Jazaieri, H., Werner, K., Kraemer, H., Heimberg, R. G., & Gross, J. J.

(2012). Cognitive reappraisal self-efficacy mediates the effects of individual cognitive-behavioral therapy for social anxiety disorder. *Journal of Consulting and Clinical Psychology, 80,* 1034-1040. doi:10.1037/a0028555

Goossens, L., Sunaert, S., Peeters, R., Griez, E. J. L., Schruers, K. R. J. (2007). Amygdala hyperfunction in phobic fear normalizes after exposure. *Biological Psychiatry, 62,* 1119-1125. doi:10.1016/j.biopsych.2007.04.024

Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014). A review on human reinstatement studies: An overview and methodological challenges. *Learning & Memory, 21,* 424-440. doi:10.1101/lm.036053.114

Haesen, K., & Vervliet, B. (2015). Beyond extinction: Habituation eliminates conditioned skin conductance across contexts. *International Journal of Psychophysiology, 98,* 529-534. doi:10.1016/j.ijpsycho.2014.11.010

Hauner, K. K., Mineka, S., Voss, J. L., & Paller, K. A. (2012). Exposure therapy triggers lasting reorganization of neural fear processing. *PNAS: Proceedings of the National Academy of Sciences of the United of America, 109,* 9203-9208. doi:10.1073/pnas.1205242109

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2003). *Acceptance and commitment therapy: An experiential approach to behavior change.* New York, NY: Guilford Press.

Hofmann, S. G. (2008). Cognitive processes during fear acquisition and extinction in animals and humans: Implications for exposure therapy of anxiety disorder. *Clinical Psychology Review, 28,* 199-210. doi:10.1016/j.cpr.2007.04.009

Hofmann, S. G., Smits, J. A. J., Rosenfield, D., Simon, N., Otto, M. W., Meuret, A. E., … Pollack, M. H. (2013). D-cycloserine as an augmentation strategy with cognitive-behavioral therapy for social anxiety disorder. *The American Journal of Psychiatry, 170,* 751-758. doi:10.1176/appi.ajp.2013.12070974

Holmes, E. A., Craske, M. G., & Graybiel, A. M. (2014). Psychological treatments: A call for mental health science. *Nature, 511,* 287-289. doi:10.1038/511287a

Jaycox, L. H., Foa, E. B., & Moral, A. R. (1998). Influence of emotional engagement and habituation on exposure therapy for PTSD. *Journal of Consulting and Clinical Psychology, 66,* 185-192. doi:10.1037/0022-006X.66.1.185

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion, 13,* 989-996. doi:10.1037/a0033164

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Counterconditioning:

An effective technique for changing conditioned preferences. *Experimental Psychology, 58,* 31-38. doi:10.1027/1618-3169/a000063

Kimmel, H. D. (1971). *Experimental Psychopathology: Recent research and theory*. New York: Academic Press.

Kindt, M., & Soeter, M. (2013). Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. *Biological Psychology, 92,* 43-50. doi:10.1016/j.biopsycho.2011.09.016.

Kircanski, K., Lieberman, M. D., & Craske, M. G. (2012). Feelings into words: Contributions of language to exposure therapy. *Psychological Science, 23,* 1086-1091. doi:10.1177/0956797612443830

Krypotos, A.- M., Arnaudova, I., Effting, M., Kindt, M., & Beckers, T. (2015). Effects of approach-avoidance training on the ectinction and return of fear responses. *PLoS ONE, 10,* e0131581. doi:10.1371/journal.pone.0131581

Leer, A., & Engelhard, I. M. (2015). Countering fear renewal: Changes in the UCS representation generalize across contexts. *Behavior Therapy, 46,* 272-282. doi:10.1016/j.beth.2014.09.012

Lipsitz, J. D., Mannuzza, S., Klein, D. F., Ross, D. C., & Fyer, A. J. (1999). Specific phobia 10-16 years after treatment. *Depression and Anxiety, 10,* 105-111.

Lommen, M. J. 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,* 63-67. doi:10.1016/j.brat.2012.11.004

Lovibond, P. F. (2004). Cognitive processes in extinction. *Learning and Memory, 11,* 495-500. doi:10.1101/lm.79604

Luyten, L., Vansteenwegen, D., van Kuyck, K., Gabriels, L., & Nuttin, B. (2011). Contextual conditioning in rats as an animal model for generalized anxiety disorder. *Cognitive, Affective, and Behavioral Neuroscience, 11,* 228-244. doi:10.3758/s13415-011-0021-6

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorder. *American Psychologist, 61,* 10-26. doi:10.1037/003-066X.61.1.10

McConnell, B. L., & Miller, R. R. (2014). Associative accounts of recovery-from-extinction effects. *Learning and Motivation, 46,* 1-15. doi:10.1016/j.lmot.2014.01.003

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65,* 73-84. doi:10.1037/a0018121

McNally, R. J. (2007). Mechanisms of exposure therapy: How neuroscience can improve

psychological treatments for anxiety disorders. *Clinical Psychology Review, 27*, 750-759. doi:10.1016/j.cpr.2007.01.003

Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological Psychiatry, 62*, 446-454. doi:10.1016/j.biopsych.2006.10.011

Milad, M. R., Quinn, B. T., Pitman, R. K., Orr. S. P., Fischl, B., & Rauch, S. L. (2005). Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. *PNAS: Proceedings of the National Academy of Sciences of the United of America, 102*, 10706-10711. doi:10.1073/pnas.0502441102

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

Mystkowski, J. L., Craske, M. G., Echiverri, A. M., & Labus, J. S. (2006). Mental reinstatement of context and return of fear in spider-fearful participants. *Behavior Therapy, 37*, 49-60. doi:10.1016/j.beth.2005.04.001

Niles, A. N., Craske, M. G., Lieberman, M. D., & Hur, C. (2015). Affect labeling enhances exposure effectiveness for public speaking anxiety. *Behaviour Research and Therapy, 68*, 27-36. doi:10.1016/j.brat.2015.03.004

Norberg, M. M., Krystal, J. H., & Tolin, D. F. (2008). A meta-analysis of d-cycloserine and the facilitation of fear extinction and exposure therapy. *Biological Psychiatry, 63*, 1118-1126. doi:10.1016/j.biopsych.2008.01.012

Ori, R., Amos, T., Bergman, H., Soares-Weiser, K., Ipser, J. C., & Stein, D. J. (2015). Augmentation of cognitive and behavioural therapies (CBT) with d-cycloserine for anxiety and related disorders. *Cochrane Database of Systematic Reviews*, 2015(5), 1-131. doi:10.1002/14651858.CD007803.pub2

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex.* London, England: Oxford University Press.

Phelps, E. A., Delgado, M. R., Nearing, K. J., & LeDoux, J. E. (2004). Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron, 43*, 897−905.

Pittig, A., van den Berg, L., & Vervliet, B. (2016). The key role of extinction learning in anxiety disorders: Behavioral strategies to enhance exposure-based treatments. *Current Opinion in Psychiatry, 29*, 39-47. doi:10.1097/YCO.0000000000000220

Rescorla, R. A. (1973). Effects of US habituation following conditioning. *Journal of*

Comparative and Physiological Psychology, 82, 137-143.

Rodrigues, H., Figueira, I., Lopes, A., Gonçalves, R., Mendlowicz, M. V., Coutinho, E. S. F., & Ventura, P. (2014). Does d-cycloserine enhance exposure therapy for anxiety disorder in humans ? A meta-analysis. *PLoS ONE, 9*, e93519. doi:10.1371/journal.pone.0093519

Rowe, M. K., & Craske, M. G. (1998$_a$). Effects of varied-stimulus exposure training on fear reduction and return of fear. *Behaviour Research and Therapy, 36*, 719-734. doi:10.1016/S0005967(97)100017-1

Rowe, M. K., & Craske, M. G. (1998$_b$). Effects of an expanding-spaced vs massed exposure schedule on fear reduction and return of fear. *Behaviour Research and Therapy, 36,* 701-717. doi:10.1016/S0005-7967(97)10016-X

Simpson, H. B., Liebowitz, M. R., Foa, E. B., Kozak, M. J., Schmidt, A. B., Rowan, V., … Campeas, R. (2004). Post-treatment effects of exposure therapy and clomipramine in obsessive-compulsive disorder. *Depression and Anxiety, 19*, 225-233. doi:10.1002/da.20003

Sloan, D. M. (2014). Introduction: Using experimental psychopathology to advance behavior therapy. *Behavior Therapy, 45*, 589-593. doi:10.1016/j.beth.2014.03.007

Soliman, F., Glatt, C. E., Bath, K. G., Levita, L., Jones, R. M., Pattwell, S. S., … Casey, B. J. (2010). A genetic variant BDNF polymorphism alters extinction learning in both mouse and human. *Science, 327,* 863-866. doi:10.1126/science.1181886

Steinert, C., Hofmann, M., Kruse, J., & Leichsenring, F. (2014). Relapse rates after psychotherapy for depression – stable long-term effects? A meta-analysis. *Journal of Affective Disorders, 168,* 107-118. doi:10.1016/j.jad.2014.06.043

Tryon, W. (2005). Possible mechanisms for why desensitization and exposure therapy work. *Clinical Psychology Review, 25,* 67-95. doi:10.1016/j.cpr.2004.08.005

Tsao, J. C. I., Craske, M. G. (2000). Timing of treatment and return of fear: Effects of massed, uniform, and expanded spaced exposure schedules. *Behavior Therapy, 31*, 479-497. doi:10.1016/S0005-7894(00)80026-X

Van den Hout, M. (1999). Armies of idiots and idiosyncrasies: On reductions in experimental psychopathology. *Behaviour Research and Therapy, 37*, 135-145.

Van den Hout, M. A., Engelhard, I. M., & McNally, R. J. (in press). Thoughts on experimental psychopathology. *Psychopathology Review*. doi:10.5127/pr.045115

Van Meurs, B., Wiggert, N., Wicker, I., & Lissek, S. (2014). Maladaptive behavioral

consequences of conditioned fear generalization: A pronounced, yet sparsely studied feature of anxiety pathology. *Behavior Research and Therapy, 57*, 29-37. doi:10.1016/j.brat.2014.03.009

Van Uijen, S. L., van den Hout, M., & Engelhard, I. (2015). Active approach does not add to the effects of in vivo exposure. *Journal of Experimental Psychopathology, 6*, 112-125. doi:10.5127/jep.042014

Vansteenwegen, D., Vervliet, B., Hermans, D., Beckers, T., Baeyens, F., & Eelen, P. (2006). Stronger renewal in human fear conditioning when tested with an acquisition retrieval cue than with an extinction retrieval cue. *Behaviour Research and Therapy, 44*, 1717-1725. doi:10.1016/j.brat.2005.10.014

Vansteenwegen, D., Vervliet, B., Iberico, C., Baeyens, F., van den Bergh, O., & Hermans, D. (2007). The repeated confrontation with videotapes of spiders in multiple contexts attenuates renewal of fear in spider-anxious students. *Behaviour Research and Therapy, 45*, 1169-1179. doi:10.1016/j.brat.2006.08.023

Vervliet, B. (2008). Learning and memory in conditioned fear extinction: Effects of d-cycloserine. *Acta Psychologica, 127*, 601-613. doi:10.1016/j.actpsy.2007.07.001

Vervliet, B., Baeyens, F., Van den Bergh, O., & Hermans, D. (2013). Extinction, generalization, and return of fear: A critical review of renewal research in humans. *Biological Psychology, 92*, 51-59. doi:10.1016/j.biopsycho.2012.01.006

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: State of the art. *Annual Review of Clinical Psychology, 9*, 215-248. doi:10.1146/annurev-clinpsy-050212-185542

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

Vervliet, B., Kindt, M., Vansteenwegen, D., & Hermans, D. (2010). Fear generalization in humans: Impact of verbal instructions. *Behaviour Research and Therapy, 48*, 38-43. doi:10.1016/j.brat.2009.09.005

Vervliet, B., & Raes, F. (2013). Criteria of validity in experimental psychopathology: Application to models of anxiety and depression. *Psychological Medicine, 43*, 2241-2244. doi:10.1017/S0033291712002267

Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of

fear in a human differential conditioning paradigm caused by stimulus change after extinction. *Behaviour Research and Therapy, 43*, 357-371. doi:10.1016/j.brat.2004.02.005

Walker, D. L., & Davis, M. (2002). The role of amygdale glutamate receptors in fear learning, fear-potentiated startle, and extinction. *Pharmacology, Biochemistry and Behavior, 71*, 379-392.

Waters, A., LeBeau, R., & Craske, M. (in press). Experimental psychopathology: An integrative model. *Psychopathology Review.*

Zlomuzica, A., Preusser, F., Schneider, S., & Margraf, J. (2015). Increased perceived self-efficacy facilitates the extinction of fear in healthy participants. *Frontiers in Behavioral Neuroscience, 9*, 1-12. doi:10.3389/fnbeh.2015.00270

Zvolensky, M. J., Forsyth, J. P., & Johnson, K. (2013). Laboratory methods in experimental psychopathology. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 7-23). New York, NY: Oxford University Press.

Chapter 7

General discussion

General Discussion

Summary and Conclusions

In this dissertation we aimed to test strategies to optimize exposure therapy with specific attention for the strengths and limitations of the used laboratory models. We started with an introduction and evaluation of the fear conditioning model (**Chapter 1**). Subsequently, in the empirical chapters of this dissertation, a first part focused on the type and order of the stimuli used during extinction. In a first study (**Chapter 2**), we tested the effect of the typicality of the extinction stimulus on the generalization of extinction after stimulus change. In particular, after extinction participants were either instructed that the extinction stimulus was a typical exemplar of the feared category (i.e., the typical group) or that it was an atypical exemplar (i.e., the atypical group). Results indicate better generalization to a new exemplar of the feared category in the typical group compared to the atypical group. Translated to clinical exposure therapy, these results suggest that using a typical member of the feared category or emphasizing the typical character of the stimuli used during exposure might enhance the generalization to other stimuli. However, before applying this in clinical practice the effectiveness of this intervention as an add-on to exposure therapy needs to be tested in clinical trials.

In a second study (**Chapter 3**), we investigated the outcome of a hierarchical (i.e., gradually moving from less to more fear-eliciting stimuli) versus a random approach of exposure in the laboratory. We developed an extinction procedure in which we presented morphs between the danger cue (CS+) and safety cue (CS-) in either a hierarchical way (i.e., starting with the CS-, then the morph most similar to the CS-, followed by the morph most similar to that one, and so on, until reaching the CS+) or in a random order. In line with the predictions of the Inhibitory Learning Theory (ILT; Craske et al., 2008; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014), higher overall US-expectancies (a proxy for expectancy violation), higher physiological arousal and higher variability in US-expectancies were found in the random extinction group compared to the hierarchical extinction group. This supports the validity of the used procedure. However, no differences were found between the hierarchical and random approach in fear responding to the original CS+ and to a new generalization stimulus one day later. These results are in line with the results of two clinical studies on this topic (Kircanski et al., 2012; Lang & Craske, 2000) and suggest that there is no reason for clinicians to refrain from using a hierarchical approach of exposure if they are for instance convinced that this would result in higher treatment acceptance and adherence.

In the second part of the empirical work in this dissertation, we focused on the role of expectancy violation, the driving mechanism underlying the formation of inhibitory associations according to ILT. In Study 3 (**Chapter 4**) we investigated how expectancy violation is involved in virtual reality exposure therapy (VRET). Participants fearful of public speaking were exposed to four exposure exercises in which they were encouraged to present about a topic in front of a virtual audience. After the VRET exercises, some participants were informed that the virtual audience could have reacted on their presentations (i.e., the interactive group) whereas other participants were informed that the reactions of the audience were not adjusted to their presentations (i.e., the non-interactive group). Notably, by giving this information participants in the interactive group were able to test some of their expectancies about the overt reactions of the audience whereas participants in the non-interactive group were less able to test this type of expectancies (e.g., "people in the audience will yawn in reaction to my speech"). In addition, we calculated a proportion of testable expectancies for each participant (i.e., expectancies indicated by the participant as being testable in VRET divided by his or her expectancies in real-life public speaking situations) and tested whether participants with a higher proportion of testable expectancies had more benefit from VRET. We found a reduction in fear-related outcomes from pre to post VRET, but no differences between the interactive and the non-interactive group. In addition, participants on average indicated expectancies about one's own reactions as being better testable in VRET compared to expectancies about the overt reactions of the audience or expectancies about being negatively evaluated. The correlational analyses, however, did not reveal clear-cut evidence for the proportion of testable expectancies being predictive for VRET outcome. In sum, we found a reduction in public speaking anxiety from pre to post VRET. The experimental and correlational analyses could not confirm a crucial role for expectancy violation in VRET. However, future research might benefit from a larger sample size resulting in more power.

In the fourth study (**Chapter 5**) we investigated the effect of providing safety information before extinction on contextual renewal. We departed from the clinical practice of providing objective information about the probability of a certain feared outcome (e.g., a plane crash) before exposure. Often the rationale for providing this type of information is to increase the willingness of a patient to engage in subsequent exposure. In ILT-based exposure, however, it is recommended to refrain from providing this type of information before exposure because this might interfere with the possibility to maximally violate expectancies. In an ABA-renewal paradigm, we tested whether providing information about the low probability of shock-

occurrence before the start of extinction resulted in more return of fear. Opposite to the predictions made by ILT, participants who received the safety information showed less contextual renewal than participants who did not receive the safety information. These results suggest that providing safety information is not detrimental for the outcome of exposure. However, testing this prediction in a clinical sample as well as including a long-term extinction retention test is recommended for future research.

After the empirical chapters, we reflected on the use of the extinction model to mimic clinical exposure therapy in the laboratory in a theoretical chapter (**Chapter 6**). Since we used the extinction model in three out of four empirical studies, we refer to this chapter for a critical evaluation and discussion of strengths and limitations of the used procedures. In addition, this chapter includes recommendations to improve the external validity of extinction research.

Integration of the Empirical Findings in an Evaluation of the Inhibitory Learning Theory (ILT)

A profound discussion of the limitations and implications of the empirical studies of this dissertation as well as recommendations for future research can be found in the individual manuscripts and the theoretical chapters. We therefore opted to continue this general discussion with an integration of our findings and conclusions within a broader reflection on ILT. ILT is currently one of the prevailing theories of exposure therapy and inspired most of our empirical studies. In particular, we aimed to test some of the assumptions and recommendations to conduct exposure therapy rooted in ILT. By embedding our findings in a discussion of the strengths of and challenges for ILT we aim to situate our findings in a broader theoretical context.

First, some important assets of ILT are discussed. Subsequently, we continue with observations that might constitute a challenge for ILT. Notably, some of these are merely challenging at first sight: they are not a substantial threat for the validity of ILT if a flexible interpretation of the framework is allowed or if other theoretical explanations are taken into account. Other observations are, however, more difficult to reconcile with ILT or require further research.

Strengths of ILT

In this section, some of the strengths of ILT are discussed. For a more elaborative review of (the assets of) ILT we refer to Pittig, Treanor, LeBeau, and Craske (2018), Pittig, van den Berg, and Vervliet (2015), and Weisman and Rodebaugh (2018).

First, while being grounded in the basic science of learning and memory, ILT has generated a number of predictions about strategies to optimize exposure therapy. Moreover, its strong affiliation with the fear conditioning framework and extinction research (as we discussed in **Chapter 1**) allows for the highly controlled experimental testing of these strategies. We illustrated the value of such translational research in **Study 1** (**Chapter 2**), in which we used a fear conditioning paradigm to demonstrate that providing information about the typicality of the extinction stimulus can attenuate return of fear.

A second strength of ILT is that it attaches much importance to the clinically relevant phenomenon of return of fear after successful exposure treatment (Vervliet, Craske, & Hermans, 2013). In particular, ILT distinguishes between 'fear expression throughout exposure' and 'learning', with the proposed strategies focusing on the latter (Craske et al., 2008). As a consequence, the primary outcomes in ILT-based treatment research are long-term fear reduction and generalization instead of fear reduction during treatment. To this end, paradigms that allow for testing return of fear and generalization are used such as reinstatement, renewal, stimulus change and long-term follow-ups/spontaneous recovery. This focus on long-term fear reduction is very useful since an important goal in optimizing exposure therapy in clinical practice is to maintain treatment effects in the long run.

Finally, neurobiological evidence that is compatible with the role of inhibitory learning in exposure therapy has been found. In particular, the connectivity between prefrontal regions and emotion regulation neural circuitry has found to be stronger after CBT and has found to be predictive for CBT outcome (e.g., Young et al., 2017; Young et al., 2018)[1].

Challenges for ILT

In addition to the undisputable strengths of ILT that can account for its current influence on the field of extinction and exposure, we continue this section with identifying some challenges.

[1] Arguably, this neurobiological evidence is not conclusive, because other psychological theories can be conceptualized that are compatible with this evidence as well.

The central role of expectancy violation

Based on models of associative learning, ILT predicts that expectancy violation or the mismatch between the expected and actual outcome(s) drives (inhibitory) learning (Rescorla & Wagner, 1972). Hence, the more expectancy violation during exposure, the stronger the inhibitory (CS-noUS) association and the better this inhibitory association will be able to compete with the excitatory (CS-US) association, resulting in diminished fear responding. Therefore, theoretically it makes sense to optimize exposure therapy by maximizing expectancy violation. However, in the next section we consider some challenges concerning the application of the expectancy violation approach in clinical practice as well as empirical evidence that is not consistent with this approach.

Patients often fail to report US-expectancies. A crucial role for US-expectancy violation implies that in exposure therapy, the expected aversive outcome (US) is made specific and concrete so that the patient can be exposed to those situations in which his expectancy of the US is maximal (and thus can be maximally violated). In clinical practice it can, however, be observed that not all anxiety patients are able to report such expectancies. For example, many spider phobics report that they know that spiders are not dangerous, but nevertheless are afraid of them. A possible explanation for this observation is the context-dependent retrieval of excitatory (fear) and inhibitory (safety) associations (for a similar argument see Mitchell, De Houwer, & Lovibond, 2009). In particular, a patient might retrieve the safety (CS-noUS) association while being in a safe context (e.g., in the absence of the CS; e.g., spider). He might have learned this safety association by verbal information, for example people telling him that spiders are harmless. However, once in proximity to the spider, the fear association is retrieved from memory and feared expectancies are reported. This context-dependent retrieval of associations can also explain some of the results that we found in **Study 4** (**Chapter 5**). In an ABA-renewal paradigm, participants in this study were told after acquisition that the probability of shock occurrence would be extremely small in the remainder of the experiment. Results show that despite this instruction (given in the absence of the CS), participants continued to expect the shock when being presented with the CS at the start of extinction.

In addition, ILT emphasizes the role of violating expectancies about *US*-occurrence. However, many patients report to expect outcomes that they experience as being very aversive, but it can be questioned whether these expected outcomes are USs. For a panic patient, for example, having a panic attack might be the aversive outcome that he expects when going to

the shopping mall, when doing physical exercises etc. However, is having a panic attack a US? The answer to this question depends on the definition of a US that is applied.

A US is sometimes defined as a *biologically significant* event, including death or severe illness (i.e., physical integrity), social rejection/exclusion, or violation of someone's psychological integrity (e.g., becoming crazy). According to this (narrow) definition, a panic attack itself would not be considered a US, but rather a CS that is associated with further USs such as having a heart attack (i.e., physical integrity) or losing control over one's behavior (i.e., psychological integrity). In particular, through higher-order conditioning the shopping mall (CS) elicits fear responding because it has become associated with a panic attack (higher-order CS), which, on its turn, has become aversive because it is associated with getting a heart attack (US). Notably, a significant part of patients considers the panic attack itself as their most feared outcome and not all patients are aware of any further consequences that they attach to panic attacks. More general, it can be considered a challenge for ILT that the feared outcomes reported by many patients do not fit into the definition of a US as being a biologically significant event.

From a functional viewpoint, however, it can be argued that CSs and USs are no ontological entities and that whether a stimulus functions as a CS or a US depends on the functional analysis (De Houwer, Barnes-Holmes, & Moors, 2013). In line with this functional definition, panic attacks can function as both a CS and a US. Panic attacks would function as a CS if they elicit a certain response (e.g., anxiety) because of their (presumed) relation to heart attacks. That is because the response to the panic attack is then conditional on the panic attack's (presumed) relation with the heart attack. However, the same panic attack can function as a US if its relation with another stimulus changes responding to that other stimulus. For example, a patient might react fearfully when being in a shopping mall where he previously experienced an aversive panic attack. In such case, responding to the shopping mall is conditional on its relation with the aversive panic attack (that now functions as a US). One could take this one step further and say that the panic attack does not function as a US but as a higher-order CS if the panic attack can only change responding to the other stimulus (e.g., the shopping mall) because of its relation with yet another stimulus (e.g., heart attack). From a functional perspective, all these scenarios are possible (i.e., panic attack as CS, US, and higher-order CS) and functional analysis should determine the exact stimuli to which the individual patient needs to be exposed.

In conclusion, we considered it a challenge for ILT that patients are not always able to report US-expectancies. However, this seemingly challenging observation can be explained by the context-dependent retrieval of feared associations. At the same time, it might be less feasible to set up an exposure treatment that aims for maximal expectancy violation if the patient is not able to identify his feared expectancies. In addition, it can be considered a challenge for ILT that many patients report feared outcomes that do not fit in the definition of a US as being a biologically significant event. Here, we argued that applying a functional definition on CSs and USs can provide a solution.

Virtual reality exposure therapy (VRET) is effective. As we argued in **Chapter 4** of this dissertation, the effects of VRET in anxiety disorders are difficult to explain by expectancy violation because (at least some) feared outcomes cannot occur in a virtual environment. For instance, being harmed in a plane crash is an outcome that cannot occur in VRET for flying phobia. Nevertheless, consistent with several meta-analyses we found a decrease in anxiety-related outcomes from pre to post VRET in **Study 3**. We also found that expectancies with regard to own reactions are considered as being more testable in VRET than expectancies about the (overt) reactions of the virtual audience and negative evaluation. However, we failed to find evidence for a relation between (the amount of) expectancy violation and VRET outcome. At first sight, these results suggest that mechanisms other than expectancy violation most likely can account for the effects of VRET. However, as we argue in the next section, the results might still be reconciled with the assumptions of ILT.

Evidence for habituation-based models of exposure. One of the main aspects in which ILT distinguishes itself from older habituation-based models is that it predicts that the long-term effects of exposure are not driven by fear habituation throughout exposure (e.g., Lader & Matthews, 1978; Foa & Kozak, 1986; Foa, Huppert, & Cahill, 2006). Fear habituation (as a mechanism) is typically inferred from within-session fear reduction (as a result) and some of the support for ILT comes from evidence showing that neither the amount of within-session fear reduction (e.g., Baker et al., 2010; Brown, LeBeau, Chat, & Craske, 2016) nor fear levels at the end of exposure are predictive for fear expressed at follow-up retesting (Prenoveau, Craske, Liao, & Ornitz, 2013). However, there is also contradictory evidence showing that within-session fear reduction does predict long-term treatment outcome (e.g., van Minnen & Hagenaars, 2002). Moreover, a recent meta-analysis found that within-session fear reduction

was positively correlated with treatment outcome (Rupp, Doebler, Ehring, & Vossbeck, Elsebusch, 2017).

Arguably, the effects of within-session fear reduction are not necessarily incompatible with ILT. In particular, it can be argued that within-session fear reduction can involve the violation of certain feared expectancies. In line with the concept 'fear of fear', the expected feared outcome can be, for instance, that fear levels would endlessly increase or that fear would lead to going crazy. The experience that fear decreases after some time (i.e., within-session fear reduction), can then be considered a violation of this expectancy. Given this extension of the expectancy violation concept, it becomes of course more difficult to distinguish between ILT and habituation-based models[2].

In conclusion, the effects of within-session fear reduction on exposure outcome are at first sight incompatible with the predictions made by ILT. However, we demonstrated that ILT is able to account for the effects of within-session fear reduction given that these effects are driven by expectancy violation and not merely by the repeated presentations of a stimulus (De Houwer et al., 2013).

Link with associative learning models

ILT draws credibility from the premise that its recommendations are based on (mathematical) association formation models such as the Rescorla-Wagner model (Rescorla & Wagner, 1972). However, not all predictions made by ILT are theoretically in line with Rescorla-Wagner. First, the prediction that random exposure would be more successful than hierarchical exposure is difficult to explain theoretically without including additional assumptions about how learning generalizes from one stimulus to another (see **Chapter 3**). However, ILT does not specify such assumptions. If the same set of stimuli is used, Rescorla-Wagner – without these additional assumptions – predicts that the net amount of expectancy violation or inhibitory strength would be the same, irrespective of whether these stimuli are presented in a random or hierarchical order. A second illustration of a prediction by ILT that deviates from the Rescorla-Wagner model is the one stating that greater variability in responding during exposure is related to enhanced treatment outcomes (e.g., Brown et al., 2017). According to Rescorla-Wagner, eliciting a high response on a specific trial allows to gain inhibitory strength. However,

[2] Notably, we use the term 'within-session fear reduction' to indicate the observable effect (i.e., decrease in fear) and the term 'habituation' to indicate the underlying mechanism. Since we argued that the observable effect (i.e., decrease in fear) can be indicative for both habituation and expectancy violation as underlying mechanisms, it becomes difficult to distinguish between both models based on a decrease in fear or anxiety during exposure.

repeating this (i.e., several increases and decreases in responding) will not add up to a greater inhibitory strength.

In addition, some of the recommendations made by ILT are in fact out of scope of the Rescorla-Wagner model. A first example of this is the recommendation to encourage affect labeling to enhance exposure therapy (e.g., Niles, Craske, Lieberman, & Hur, 2015). Another example is the prediction that providing safety information before exposure would interfere with the outcome of exposure (a prediction that was tested in **Study 4**). This prediction relies on the (implicit) assumption that learning via experience is mediated by a different mechanism than learning via instruction and therefore can interfere with it (i.e., dual process models). Although Rescorla-Wagner is silent about learning via instructions, other approaches of associative learning, such as the propositional approach, would actually make the exact opposite prediction and would predict an added effect of instructions (e.g., Mitchell et al., 2009).

Needless to say, the observation that not all treatment recommendations by ILT are grounded in prediction error models like Rescorla-Wagner is not a major issue for its clinical application. However, it should be noted that ILT draws some of its credibility from the link with these prediction error models. Moreover, we note that for some of the predictions that are framed as being rooted in the Rescorla-Wagner model the link with that model is not as straightforward as typically assumed.

Extinction versus exposure

As discussed in the first section, the theory-based character of ILT has generated useful recommendations to optimize exposure therapy and the availability of a laboratory model to test these recommendations is a major asset of ILT. Some of the ILT-based exposure strategies are even exclusively tested in extinction research (e.g., occasional reinforcement, deepened extinction).

The question is, however, whether findings from extinction research automatically translate to clinical practice. Although this is often assumed to be the case (e.g., Graham & Milad, 2011), we argued in **Chapter 6** that in addition to the similarities between extinction and exposure therapy there might be some crucial differences. In particular, evidence for additional mechanisms underlying the effects of clinical exposure therapy was discussed, including habituation, self-efficacy, and approach behavior. However, are these mechanisms necessarily incompatible with inhibitory learning? In an earlier section we suggested that

within-session fear reduction that is typically considered evidence for the habituation account can involve the violation of certain expectancies. We now develop similar arguments with regard to self-efficacy and approach behavior.

First, in exposure therapy patients might acquire a sense of self-efficacy or perceived ability that one is able to cope with the feared situation (Bandura, 1997). We will now reflect on how expectancy violation and inhibitory learning can account for this. Before exposure, the patient might belief that he is not able to cope with the feared situation. However, 'not being able to cope with the feared situation' can be reframed in specific expectancies about a confrontation with the feared situation (e.g., by asking questions like "How can we observe that you are not able to cope with the feared situation?"). In flying phobia for example: "I will become crazy and lose control over my behavior, I will open the door of the plane while we are flying". These expectancies can subsequently be tested and violated in exposure therapy.

Second, approaching the feared situation or stimulus has been proposed as an active component in exposure therapy (i.e., fear antagonistic actions; Wolitzky & Telch, 2009). The effects of approach behavior can be compatible with ILT depending on the exact underlying working mechanisms that are assumed to account for its effects. First, new Stimulus-Response (SR) associations can be formed by approaching the feared stimulus. This explanation is not compatible with ILT if ILT is considered to be a Stimulus-Stimulus model in which associations between stimuli are learned (i.e., CS-US and CS-noUS)[3]. Second, fear antagonistic actions or approach behavior (e.g., asking an acrophobic patient to run towards the rail of the balcony) can be considered as directly opposite to and incompatible with certain safety behaviors and therefore allow for more expectancy violation (Weisman & Rodebaugh, 2018). Evidently, this explanation is compatible with ILT.

In conclusion, in translating findings from extinction research to clinical exposure treatment, it is important to recognize the differences between extinction and exposure. Notably, other mechanisms in addition to inhibitory learning are proposed to be at work in exposure therapy. We illustrated that some of these alternative mechanisms are, however, not necessarily incompatible with ILT.

[3] Notably, this is not explicitly specified in ILT. However, the notions CS-US and CS-noUS associations that are used in the theoretical papers on ILT suggest that it relates to SS-learning.

Clinical application of ILT-based exposure

In this section, we dwell on possible challenges concerning the clinical application of ILT-based exposure.

ILT-based exposure is a demanding approach of exposure. Compared to more traditional approaches of exposure, some of the recommendations to maximize inhibitory learning during exposure might require an extra effort and a willingness to tolerate high levels of distress from the patient who seeks treatment.

First, based on ILT it is recommended to encourage patients to eliminate safety behaviors or behaviors that are perceived to neutralize threat or reduce anxiety during exposure, because these behaviors could interfere with the opportunity to maximally violate expectancies (e.g., Blakey & Abramowitz, 2016). However, empirical evidence for this recommendation is limited. A recent meta-analytic review found no compelling evidence for larger effect sizes in exposure without safety behaviors compared to exposure in which safety behaviors were allowed (Meulders, Van Daele, Volders, & Vlaeyen, 2016; see also Milosevic & Radomsky, 2008). Similar results were found for attentional distraction during exposure, which can also be considered a safety behavior. A recent meta-analysis by Podina, Koster, Philippot, Dethier, and David (2013) could not confirm that attentional distraction was detrimental for the outcome of exposure treatment. At the same time, it has been demonstrated that the judicious use of safety behaviors can enhance the acceptability of exposure (Levy & Radomsky, 2014, but see Deacon, Sy, Lickel, & Nelson, 2010).

Second, providing cognitive restructuring and safety information before exposure is discouraged in ILT-based exposure (Weisman & Rodebaugh, 2018). As we discussed in **Chapter 5** and similar to the recommendation to refrain from safety behaviors, providing this type of information could interfere with expectancy violation during exposure. However, this recommendation has never been tested empirically in a clinical sample. Moreover, we failed to find evidence for deleterious effects of providing safety information before extinction in a fear conditioning paradigm (**Study 4**).

Third, it is recommended by ILT to conduct exposure in a random way (i.e., exposure to less and more fear-eliciting items randomly) instead of in a hierarchical way (i.e., gradually proceeding from less to more fear-eliciting stimuli). It is hypothesized that random exposure is less predictable, produces more sustained arousal, higher levels of variability and higher

expectancy violation. Using a fear conditioning paradigm we confirmed in **Study 2** (**Chapter 3**) that random extinction resulted in higher sustained arousal, higher variability in US-expectancies and overall higher US-expectancies throughout extinction. However, we did not find evidence for less fear responding at retest in random extinction compared to hierarchical extinction. These results are consistent with two clinical studies comparing random and hierarchical exposure treatment (Kircanski et al., 2012; Lang & Craske, 2000). Hence, this recommendation is not supported by empirical evidence. Moreover, experiencing within-session fear reduction can be very reinforcing for patients and a lack of within-session fear reduction might be associated with treatment discontinuation (e.g., Norton, Hayes-Skelton, & Klenck, 2011).

In conclusion, recommendations grounded in ILT aim to maximize inhibitory learning by maximizing expectancy violation during exposure. Behavior and interventions that are thought to reduce expectancies of the aversive outcome and as a consequence most likely also reduce distress, are discouraged in ILT-based exposure. Therefore, this approach most likely requires more effort from patients in tolerating high levels of distress compared to more traditional approaches of exposure, with the potentially adverse side effect that it might cause more patients to refuse or dropout from treatment. At the same time, the empirical evidence for (at least some of) these recommendations is limited. Further research is therefore needed to test whether the effects of applying these recommendations outweigh potential higher treatment refusal and dropout.

What is needed to consider ILT-based exposure an evidence-based treatment? The potential of ILT to translate theory-based predictions and recommendations to clinical practice can be considered one of its major strengths (cf. supra; e.g., Tolin, 2018). However, given that no RCTs are published that directly compare an ILT-based approach of exposure (including the complete set of treatment recommendations) with traditional exposure, it can be argued that the empirical support of ILT-based exposure is currently insufficient to consider it as an evidence-based treatment (e.g., Tolin, McKay, Forman, Klonsky, & Thombs, 2015)[4]. Most recommendations or strategies are studied separately and not all of these individual recommendations are equally well supported by empirical evidence. Moreover, some of the proposed strategies are only tested in extinction research (cf. supra). In sum, more clinical

[4] Notably, Michelle Craske and her team are currently conducting such RCT at UCLA.

research and RCTs are needed to provide a solid empirical base for the larger-scale dissemination of ILT-based exposure principles.

Conclusion

In this general discussion, we first summarized the main conclusions of this dissertation and some implications for exposure therapy. Most of the empirical work in this dissertation was inspired by predictions and assumptions of a prevailing theory of exposure therapy (i.e., the Inhibitory Learning Theory), but our findings were not always consistent with these predictions. We continued this general discussion with an integration of the findings of the empirical studies in a broader reflection on some of the strengths of and challenges for ILT.

As illustrated by the empirical work in this dissertation, the theory-based character of ILT has contributed to the generation of strategies to optimize exposure therapy and its close link with fear conditioning offers the unique possibility to test these recommendations in the laboratory. Although science-based, ILT-based exposure is currently not evidence-based yet. Further research is needed that supports the effects of the proposed strategies in clinical samples and the need for large scale RCTs comparing ILT-based exposure with traditional exposure is pressing.

References

Baker, A., Mystkowski, J., Culver, N., Yi, R., Mortazavi, A., & Craske, M. G. (2010). Does habituation matter? Emotional Processing Theory and exposure therapy for acrophobia. *Behaviour Research and Therapy, 48,* 1139-1143. doi:10.1016/j.brat.2010.07.009

Bandura, A. (1977). *Self-efficacy: The exercise of control.* New York, NY: Freeman.

Blakey, S. M., & Abramowitz, J. S. (2016). The effects of safety behaviors during exposure therapy for anxiety: Critical analysis from an inhibitory learning perspective. *Clinical Psychology Review, 49,* 1-15. doi:10.1016/j.cpr.2016.07.002

Brown, L., LeBeau, R. T., Chat, I., & Craske, M. G. (2016). Associative learning versus fear habituation as predictors of long-term extinction retention. *Cognition and Emotion, 31*, 687-698. doi:10.1080/02699931.2016.1158695

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, N., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy, 46,* 5-27. doi: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. doi:10.1016/j.brat.2014.04.006

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychological Bulletin Review, 20,* 631-642. doi:10.3758/s13423-013-0386-3

Deacon, B. J., Sy, J. T., Lickel, J. J., & Nelson, E. A. (2010). Does the judicious use of safety behaviors improve the efficacy and acceptability of exposure therapy for claustrophobic fear? *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 71-80. doi:10.1016/j.jbtep.2009.10.004

Foa, E. B., Huppert, J. D., Cahill, S. P. (2006). Emotional Processing Theory: An update. In B. O. Rothbaum (Ed.), *Pathological anxiety: Emotional processing in etiology and treatment* (pp. 3-24). New York, NY, US: Guilford Press.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin, 99,* 20-35. doi:10.1037/0033-2909.1.20

Kircanski, K., Mortazavi, A., Castriotta, N., Baker, A. S., Mystkowski, J. L., Yi, R., & Craske, M. G. (2012). Challenges to the traditional exposure paradigm: Variability in exposure therapy for contamination fears. *Journal of Behavior Therapy and*

Experimental Psychiatry, 43, 745-751. doi:10.1016/j.jbtep.2011.10.010

Lader, M. H., & Mathews, A. M. (1968). A physiological model of phobic anxiety and desensitization. *Behaviour Research and Therapy, 6*, 411-421. doi:10.1016/0005-7967(68)90021-1

Lang, A. J., & Craske, M. G. (2000). Manipulations of exposure-based therapy to reduce return of fear: A replication. *Behaviour Research and Therapy, 38*, 1-12. doi:10.1016/S0005-7967(99)00031-5

Levy, H. C., & Radomsky, A. S. (2014). Safety behaviour enhances the acceptability of exposure. *Cognitive Behaviour Therapy, 43*, 83-92. doi:1080/16506073.2013.819376

Meulders, A., Van Daele, T., Volders, S., & Vlaeyen, J. W. S. (2016). The use of safety-seeking behavior in exposure-based treatments for fear and anxiety: Benefit of burden? A meta-analytic review. *Clinical Psychology Review, 45*, 144-156. doi:10.1016/j.cpr.2016.02.002

Milosevic, I., & Radomsky, A. S. (2008). Safety behaviour does not necessarily interfere with exposure therapy. *Behaviour Research and Therapy*, *46*, 1111-1118. doi:10.1016/j.brat.2008.05.011

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183-198.

Norton, P. J., Hayes-Skelton, S. A., Klenck, S. C. (2011). What happens in session does not stay in session: Changes within exposure predict subsequent improvement and dropout. *Journal of Anxiety Disorders, 25*, 654-660. doi:10.1016/j.janxdis.2011.02.006

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

Pittig, A., van den Berg, L., & Vervliet, B. (2015). The key role of extinction learning in anxiety disorders: Behavioral strategies to enhance exposure-based treatment. *Current Opinion in Psychiatry, 28*. doi:10.1097/YCO.0000000000000220

Podina, I. R., Koster, E. H. W., Philippot, P., Dethier, V., & David, D. O. (2013). Optimal attentional focus during exposure in specific phobia: A meta-analysis. *Clinical Psychology Review, 33*, 1172-1183. doi:10.1016/j.cpr.2013.10.002

Prenoveau, J. M., Craske, M. G., Liao, B., Ornitz, E. M. (2013). Human fear conditioning and extinction : Timing is everything… or is it ? *Biological Psychology*, *92*, 59-68. doi:10.1016/j.biopsycho.2012.02.005

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

Rupp, C., Doebler, P., Ehring, T., & Vossbeck-Elsebusch, A. N. (2017). Emotional Processing Theory put to test: A meta-analysis on the association between process and outcome measures in exposure therapy. *Clinical Psychology and Psychotherapy, 24*, 697-711. doi:10.1002/cpp.2039

Tolin, D. F. (2018). Inhibitory learning for anxiety-related disorders. *Cognitive and Behavioral Practice*.

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. *Clinical Psychology: Science and Practice, 22*, 317-338. doi:10.111/cpsp.12122

van Minnen, A., & Hagenaars, M. (2002). Fear activation and habituation patterns as early process predictors of response to prolonged exposure treatment in PTSD. *Journal of Traumatic Stress, 15*, 359-367. doi:10.1023/A:1020177023209

Vervliet, B., Craske, M. G., & Hermans, D. (2013): Fear extinction and relapse: State of the art. *Annual Review of Clinical Psychology, 9*, 215-248. doi:10.1146/annurev-clinpsy-050212-185542

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. doi:10.1016/j.cpr.2017.10.010

Wolitzky, K. B., & Telch, M. J. (2009). Augmenting in vivo exposure with fear antagonistic actions: A preliminary test. *Behavior Therapy, 40*, 57-71.

Young, K. S., Burklund, L. J., Torre, J. B., Saxbe, D., Lieberman, M. D., & Craske, M. G. (2017). Treatment for social anxiety disorder alters functional connectivity in emotion regulation circuitry. *Psychiatry Research Neuroimaging, 261*, 44-51.

Young, K. S., LeBeau, R. T., Niles, A. N., Hsu, K. J., Burklund, L. J., Mesri, B.,…, & Craske, M. G. (2018). Neural connectivity during affect labeling predicts treatment response to psychological therapies for social anxiety disorder.