

**A review on the effects of verbal instructions in human fear conditioning: Empirical findings, theoretical considerations, and future directions**

Gaëtan Mertens<sup>1,2</sup>, Yannick Boddez<sup>3,4</sup>, Dieuwke Sevenster<sup>1</sup>, Iris M. Engelhard<sup>1</sup>, & Jan De Houwer<sup>2</sup>

<sup>1</sup>Department of Clinical Psychology, Utrecht University, Utrecht, the Netherlands

<sup>2</sup>Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

<sup>3</sup>Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Groningen, The Netherlands

<sup>4</sup>Centre for the Psychology of Learning and Experimental Psychopathology, KU Leuven, Leuven, Belgium

*In press. Biological Psychology.*

Correspondence concerning this article should be addressed to Gaëtan Mertens, Department of Clinical Psychology, Heidelberglaan 1, 3584CS, Utrecht University, Utrecht, the Netherlands.

E-mail: g.mertens@uu.nl

Tel: +31 6 235 311 80

### Abstract

Fear learning reflects the adaptive ability to learn to anticipate aversive events and to display preparatory fear reactions based on prior experiences. Usually, these learning experiences are modeled in the lab with pairings between neutral conditioned stimuli and aversive unconditioned stimulus (i.e., fear conditioning via CS-US pairings). Nevertheless, for humans, fear learning can also be based on verbal instructions. In this review, we consider the role of verbal instructions in laboratory fear learning. Specifically, we consider both the effects of verbal instructions on fear responses in the absence of CS-US pairings as well as the way in which verbal instructions moderate fear established via CS-US pairings. We first focus on the available empirical findings about both types of effects. More specifically, we consider how these effects are moderated by elements of the fear conditioning procedure (i.e., the stimuli, the outcome measures, the relationship between the stimuli, the participants, and the broader context). Thereafter, we discuss how well different mental-process models of fear learning account for these empirical findings. Finally, we conclude the review with a discussion of open questions and opportunities for future research.

Keywords: Instructions; Learning; Fear; Conditioning;

### Highlights

- We provide a review of the effects of verbal instructions in human fear conditioning.
- We consider the effects CS-US contingency instructions, without any CS-US pairings.
- We also consider how verbal instructions modulate the effects of CS-US pairings.
- The implications of this research for theories of fear learning are discussed.
- We identify gaps in the literature and outline directions for future research.

## 1. Introduction

Fear conditioning is a relatively simple procedure that is often used in different research areas such as experimental psychopathology, animal behavior, behavioral neuroscience, and psychopharmacology. In this procedure, a conditioned stimulus, CS, is repeatedly paired with an aversive unconditioned stimulus, US, which results in the establishment of conditioned fear responses to the CS. In humans, this procedure is usually implemented by pairing a visual CS (e.g., a blue square) with a mildly aversive US (e.g., a calibrated electric shock), and presenting another CS (e.g., a yellow square) that is not followed by the shock. As a result of this acquisition phase, the first CS (referred to as the CS+) typically evokes more fear than the second CS (referred to as the CS-). The conditioned fear response is assumed to involve subjective, physiological, and behavioral components that can be assessed using self-reports (e.g., of subjective fear or US expectancy), physiological responses (e.g., skin conductance), and behavioral responses (e.g., approach-avoid responses such as pressing a button to avoid the US). The fear acquisition phase can be supplemented with other phases, for instance, an extinction phase during which the CS+ is presented in the absence of the US. Many factors have been varied in fear conditioning research, including the type of CSs (e.g., evolutionary relevant stimuli, such as pictures of snakes), the context in which stimuli are presented (e.g., the color of a background screen), or the type of population (e.g., anxiety patients; see Lonsdorf et al., 2017, for an extensive overview of relevant factors in fear conditioning). Fear conditioning research is important because it provides insight into the adaptive capacity of humans and other animals to learn which cues predict the occurrence of aversive and potentially dangerous events (that is, the fear conditioning procedure provides an important insight into the process of fear learning; for a further clarification regarding the difference between procedures and processes see LeDoux,

2014).

In the current review, we will address the role of verbal instructions in human fear conditioning. On the one hand, we review evidence about the effects of conditioning instructions, that is, instructions about CS-US relations (e.g., telling people that a blue square will be followed by a mild shock). On the other hand, we consider the way in which instructions about various elements of a conditioning procedure (e.g., the nature of the CSs) moderate the impact of actual CS-US pairings on conditioned fear responses (see below for more information about the specific procedural elements that we will focus on). Understanding the effects of instructions in human fear conditioning research is important for a number of reasons. First, it has already been known for a long time that verbal instructions about CS-US contingencies can result in fear for the CS (Cook & Harris, 1937). However, this capacity of verbal instructions about CS-US contingencies to install fear has only received little consideration in the fear conditioning literature, until recently. Second, verbal instructions can moderate the effects of CS-US pairings on fear. They might even influence the nature of the processes via which CS-US pairings lead to fear (Lonsdorf et al., 2017). Third, from a practical and ethical point of view, it is not possible to avoid instructions altogether in fear conditioning research with humans, because participants have to be informed about the procedures to know what is expected of them and to provide informed consent to participate in the studies. Given these considerations, we think it is important to further our understanding of the effects of verbal instructions in human fear conditioning and the implications of these studies on theories about fear learning. Such a review has in part been undertaken before (Field, 2006; Fuhrer & Baer, 1969; Grings, 1973; Luck & Lipp, 2016a; Muris & Field, 2010). However, these reviews did not attempt to provide a comprehensive overview of the effects of verbal instructions in fear conditioning but mostly focused on specific topics (such

as the role of instructions on the extinction of conditioned fear, Luck & Lipp, 2016a, or the effects of verbal threatening instructions in children, Muris & Field, 2010). Moreover, during the last few years, there has been a stark increase in the number of research articles focusing on this topic. Therefore, we think that an updated and more comprehensive review of studies investigating the role of verbal instructions in fear conditioning is due. In this paper, we aim to provide such a review.

To organize the research on the effects of verbal instructions in fear conditioning, we identify five core procedural elements of fear conditioning procedures (for papers that use a similar framework see: De Houwer, 2011; Lipp, 2006; Lonsdorf et al., 2017). Any conditioning experiment involves pairing a CS and a US that elicits a specific unconditioned reaction (UR). These pairings between the CS and the US result in the establishment of conditioned responses (CR) to the CS. This description highlights the three first core elements of the conditioning procedure: The stimuli (CS and US), the outcome measures (UR and CR) and the relationship between the stimuli (e.g., the number of pairings, the statistical contingency between the stimuli, the temporal relationship between the stimuli). Furthermore, these pairings are not administered in a void, but are presented to a specific participant (with certain characteristics) in a broader context with certain task demands and distractors. We will use these five elements of a conditioning procedure (the stimuli used, the outcome measures, the relationship between stimuli, the characteristics of the participant and the distractors and task demands of the broader context) to discuss both the effects of instructions when there are no CS-US pairings as well as the way in which verbal instructions moderate the effects of CS-US pairings on fear. In our review, we only include studies that: (1) used an aversive US (or verbally implied the presence of such a US; see Section 2.1.3), (2) provided explicit instructions about one of the elements of a fear conditioning

procedure, and (3) measured one or more behavioral or physiological outcome measure of conditioned fear (see Section 2.2). As such, we will not consider studies that exclusively deal with non-aversive USs and include only measures of liking or contingency ratings (i.e., studies that exclusively deal with contingency or evaluative learning). Furthermore, we will not consider subtle instructional effects of procedural elements other than direct verbal instructions (e.g., the possibility that participants might experience the request to rate their expectancy of the US as an instruction to learn about CS-US contingencies). Finally, in order to limit the scope of our review, we do not consider the effects of instructions on neural activity in brain regions such as the amygdala, and the cingulate and insular cortex (e.g., Büchel, Morris, Dolan, & Friston, 1998; see Mechias, Etkin, & Kalisch, 2010, for a review and meta-analysis of that literature).

After considering the procedural knowledge (i.e., the way in which effects depend on specific elements of the procedure) about the effect of verbal instructions on fear conditioning, we discuss how these effects relate to several mental-process theories that attempt to explain how fear learning takes place. Specifically, we consider theories that propose that learned fear is the result of conscious expectations about the presence of an aversive event in the presence of certain antecedent stimuli (Davey, 1992; Lovibond, 2011; Reiss, 1980) and compare those with theories that propose that learned fear is the result of automatic associative learning processes (LeDoux, 2014; Öhman & Mineka, 2001; Olsson & Phelps, 2007). Finally, we conclude our paper with an overview of open questions and avenues for future research.

## 2. Effects of Verbal Instructions in the Absence of CS-US Pairings: Fear Conditioning via Verbal Instructions<sup>1</sup>

Ever since the work by Cook and Harris (1937), we know that verbal instructions about the contingency between a CS and an aversive US (e.g., “this green light will be followed by an electric shock”), in the absence of any actual CS-US pairings, can result in conditioned fear responses towards the CS. In line with the original terminology of Cook and Harris, we refer to this procedure (i.e., verbally specifying a spatio-temporal contingency between a CS and a US) as ‘fear conditioning via verbal instructions’<sup>2</sup>. Here, we review the effects of the stimuli, the outcome measures, the relationship between stimuli, the characteristics of the participant and the broader context on fear conditioning via verbal instructions. An overview of this section can be found in Table 1.

### 2.1. Effects of stimuli

#### 2.1.1. Type of CS

Fear conditioning via verbal instructions has been demonstrated with various types of visual conditioned stimuli, such as geometric shapes (Costa, Bradley, & Lang, 2015; Mertens &

---

<sup>1</sup> Note that we will specify from here on which specific procedure was used to install conditioned fear: verbal instructions (i.e., verbally specifying a contingency between a CS and a US, or implying such a contingency, in the absence of any CS-US pairings, see Section 2.1.3) or CS-US pairings.

<sup>2</sup> In a strict sense, ‘fear conditioning via verbal instructions’ might be an inappropriate usage of the term conditioning. Conditioning refers to the effects of the spatio-temporal pairing of (conditioned and unconditioned) stimuli. Hence, unless it is assumed that the co-occurrence of words referring to a conditioned stimulus and an unconditioned stimulus in a sentence constitutes a stimulus pairing and that the effects of the verbal instructions are due to this spatio-temporal pairing of words (see Field, 2006; Gast & De Houwer, 2012 for such an argument), it seems inappropriate to talk about ‘fear conditioning via verbal instructions’. Rather, it seems likely that verbal instructions produce their effects because of their symbolic meaning rather than because of their spatio-temporal properties (see De Houwer & Hughes, 2016). Thus, strictly speaking it would be more correct to talk about ‘the effects of conditioning instructions on fear’. However, because ‘fear conditioning via verbal instructions’ is more common terminology and because the meaning of this terminology is generally clear, we decided to use this more conventional phrasing (see De Houwer & Hughes, 2016, for a more extended discussion).



De Houwer, 2016a), colored lights (Cook & Harris, 1937; Grillon, Ameli, Merikangas, Woods, & Davis, 1993), pictures of unknown animals (Field & Storksen-Coulson, 2007; Ugland, Dyson, & Field, 2013), pictures of faces (Olsson & Phelps, 2004) and nonsense words (Bennett, Vervoort, Boddez, Hermans, & Baeyens, 2015). These studies indicate that fear conditioning via verbal instructions is a quite general phenomenon. However, to our knowledge, fear conditioning via verbal instructions has not been demonstrated with stimuli in other sensory modalities, such as auditory or tactile stimuli.

### **2.1.2. Stimulus preparedness**

One specific case concerns stimuli that are thought to be evolutionary prepared to elicit fear (Öhman & Mineka, 2001; Seligman, 1971). Previous studies using CS-US pairings to establish conditioned fear have found that fear is acquired more rapidly and is more resistant to extinction (i.e., the unpaired presentation of the CS, which usually results in a reduction of conditioned fear responses) when biologically fear-relevant stimuli, such as pictures of spiders and snakes, are used as CSs than when fear-irrelevant stimuli are used, such as pictures of butterflies or flowers (Öhman & Mineka, 2001). Interestingly, similar findings have been observed for fear conditioning via verbal instructions: Instructing participants that pictures of spiders and snakes will be followed by an electric shock leads to stronger conditioned fear responses (as measured with skin conductance responses and subjective ratings, see Section 2.2) (Hugdahl & Öhman, 1977; Mertens, Raes, & De Houwer, 2016) and resistance to instructed extinction (see Section 3.3.2) (Hugdahl, 1978) than when participants are instructed that pictures of birds or butterflies will be followed by an electric shock. Thus, evolutionary preparedness of the CSs appears to play a role in fear conditioning via verbal instructions as well.

### **2.1.3. Type of US**

Most of the studies on fear conditioning via verbal instructions involved telling participants that a certain neutral stimulus would be followed by an electric shock. However, conditioned fear via verbal instructions can also be produced by providing verbal information about the delivery of either loud sounds, air blasts to the throat or a compound of a loud sound and an unpleasant image in the presence of a certain neutral CS (Bennett et al., 2015; Grillon & Ameli, 1998).

In some cases, the US is not explicitly described but merely implied by the nature of threatening instructions about the CS (e.g., Field & Lawson, 2003; Field & Storksen-Coulson, 2007; Ugland & Field, 2013). For instance, participants might be informed that a certain animal has long claws, sharp teeth, carries diseases, and is feared and avoided by people (see Table 2 for an example). Such threatening instructions differ from the conditioning via verbal instructions procedure because the US and the spatio-temporal contingency between the CS and a US are not described. However, the threatening instructions *imply* the presence of a US and a CS-US contingency (e.g., this animal might bite and infect me). Hence, effects of threatening information on fear responses might be seen as a specific case of fear conditioning via verbal instructions where a US and CS-US contingency is implied, rather than explicitly stated.

### **2.1.4. Stimulus generalization**

Effects of learning are usually not limited to the specific stimuli used within the learning situation, but tend to generalize to perceptually and conceptually similar stimuli (Dougher, Augustson, Markham, Greenway, & Wulfert, 1994; Dunsmoor, Martin, & LaBar, 2012; Lissek et al., 2008; Struyf, Zaman, Vervliet, & Van Diest, 2015). Generalization effects have also been

demonstrated for fear conditioning via verbal instructions. For instance, Bublatzky and Schupp (2012) told participants that either “pleasant”, “neutral” or “unpleasant” pictures would get paired with an electric shock. Participants were then shown 60 pictures of the International Affective Picture System (Lang, Bradley, & Cuthbert, 1997) of which 20 pictures had pleasant content, 20 pictures had neutral content and 20 pictures had unpleasant content. Although no actual shocks were administered, the results of this experiment indicated that the effects of the shock instructions (as indicated by shock expectancy ratings and the amplitudes of event-related potentials components) generalized to all the exemplars within the conceptual (pleasant, neutral or unpleasant pictures) category that was threatened. That is, the effects of threat instructions were observed for the exemplars within a category, although none of these exemplars were actually paired with an electric shock.

In another study by Bennett, Vervoort, Boddez, Hermans and Baeyens (2015), participants were trained to conceptually relate stimuli (sounds, nonsense words and nonsense figures) through a matching-to-sample (MTS) training task. In a MTS task, participants are rewarded to select one stimulus in the presence of another stimulus and thus learn to relate these stimuli (for a more extensive introduction to this procedure see Hermans & Baeyens, 2013). After this training, one of the stimuli from one trained conceptual class was paired with verbal threatening information (i.e., presented together with the verb “is” and the nouns “injury”, “terrible”, “danger”, “pain” and “hurt”), while a stimulus from another conceptual class was paired with positive information (i.e., presented together with the verb “is” and the nouns “safe”, “secure”, “gentle”, “trust” and “peace”). Bennett et al. (2015) found that the effects of threatening information (as measured by avoidance responses, US expectancy ratings, and CS valence ratings) generalized to all stimuli within the conceptual class that included the threatened

stimulus. Furthermore, the effects of threat instructions generalized to a stimulus that was perceptually similar to one of the stimuli of the conceptual class, but that was not included in the MTS training task. Hence, these studies of Bublatzky and Schupp (2012) and Bennett et al. (2015) demonstrate that fear responses installed via verbal instructions can generalize to conceptually and perceptually related stimuli as well.

#### **2.1.5. Perceptibility of the CSs**

Several experiments have investigated whether conditioned fears can be observed even under conditions that seemingly preclude conscious perception of stimuli (e.g., when presenting one stimulus briefly and another image right afterwards to perceptually ‘mask’ the first stimulus; Marcel, 1983). Olsson and Phelps (2004) used such a masking procedure to compare fear responses (as measured with skin conductance responses) that resulted from prior fear conditioning via stimulus pairings, verbal instructions or social observation (i.e., watching a movie clip of a person receiving an electric shock in the presence of the CS). They found comparable expression of conditioned fear in all three learning conditions when CSs were clearly visible. However, when CSs were masked, conditioned fear was observed only with stimulus pairings and social observation, but not with verbal instructions. This suggests that fear established via verbal instructions may be less robust when stimulus perceptibility is reduced compared to fear established via stimulus pairings or via social observation.

Table 1. Overview of the available published articles investigating the effects of different variations of procedural elements on fear conditioning via verbal instructions.

<b>Procedural elements</b>		<b>Fear conditioning via CS-US pairings</b>	<b>Fear conditioning via verbal instructions</b>
<b>Stimuli</b>	Fear conditioning with visual CSs	Switzer (1933)	Cook & Harris (1937) Grillon et al. (1993) Olsson & Phelps (2004)
	Fear conditioning with biologically prepared CSs	Öhman & Mineka (2001)	Hugdahl & Öhman (1977) Hugdahl (1978) Mertens, Raes, et al. (2016)
	Fear conditioning with exteroceptive USs	Freeman (1930)	Cook & Harris (1937) Grillon et al. (1993) Costa et al. (2015)
	Generalization to perceptually similar stimuli	Lissek et al. (2008)	Bennett et al. (2015)
	Generalization to conceptually similar stimuli	Dougher et al. (1994)	Bublitzky & Schupp (2012) Bennett et al. (2015)
	Conditioned fear responses to masked CSs	Esteves et al. (1994)	Olsson & Phelps (2004)
<b>Outcome measures</b>	Self-reported fear	Hermans et al. (2005)	Field & Lawson (2003) Soeter & Kindt (2012) Raes et al. (2014)
	US expectancy ratings	Dawson & Biferno (1973)	Raes et al. (2014) Mertens, Kuhn, et al. (2016) Mertens & De Houwer (2016a; 2016b)
	CS valence ratings	Hermans et al. (2002)	Bennett et al. (2015)
	Reasoning biases	Tomarken & Mineka (1989)	Remmerswaal et al. (2014)
	Skin conductance	Dawson & Furedy (1976)	Cook & Harris (1937) Olsson & Phelps (2004) Costa et al. (2015)
	Heart rate	Lipp & Vaitl (1990)	Costa et al. (2015)
	Facial EMG	Dimberg (1987)	Costa et al. (2015)
	Startle reflex	Spence & Runquist (1958)	Grillon & Ameli (1998) Costa et al. (2015) Mertens & De Houwer (2016a; 2016b)
	fMRI activation	Büchel et al. (1998)	Phelps et al. (2001)

	Event-related potentials	Miskovic & Keil (2012)	Bublitzky & Schupp (2012) Weymar et al. (2013)
	Avoidance behavior	Lovibond et al. (2009)	Dymond et al. (2012) Bennett et al. (2015)
	Visual attention bias	Koster et al. (2005)	Deltomme et al. (2017)
	Implicit measures of evaluation	Hermans et al. (2002)	Field & Lawson (2003)
<b>Variations of the regularity</b>	Extinction	Hermans et al. (2006)	Hugdahl et al. (1977) Mertens & De Houwer (2016b)
	Reinstatement	Haaker et al. (2014)	Mertens, Raes, et al. (2016) Mertens, Kuhn, et al. (2016)
	Renewal	Vervliet et al. (2013)	Mertens & De Houwer (2016b)
	Stimulus competition	Hinchy et al. (1995)	Lovibond (2003)
<b>Characteristics of the participants</b>	Trait anxiety	Grillon et al. (2002)	Grillon et al. (1993)
	Early age	Gao et al. (2010)	Field & Lawson (2003) Muris & Field (2010)
<b>Elements of the broader context</b>	No available studies	NA	NA

---

Note: For comparison purposes we have provided a reference to a related article on fear conditioning via CS-US pairings. Further note that it is beyond the scope of our paper to provide an exhaustive overview of the role of procedural elements in fear conditioning procedures. For a recent in-depth overview of the factors that influence fear conditioning, see Lonsdorf et al. (2017).

## 2.2. Outcome measures

Different outcome measures of conditioned fear are often classified according to Lang's fear response systems (Lang, 1968): subjective, physiological and behavioral fear responses. We use this classification to discuss whether instructions can install verbally conditioned fear responses in all these response systems.

### **2.2.1. Subjective measures**

Subjective measures of conditioned fear often include ratings of distress, fear, CS valence and expectancy of the US (e.g., Boddez et al., 2012; Lipp, 2006). These measures are typically collected with visual analog scales or Likert scales online (i.e., during every CS presentation) or retrospectively (i.e., at the end of the experiment). Several studies have found that fear conditioning via verbal instructions produces increased ratings of self-reported distress, fear, CS unpleasantness and US expectancy (Bennett et al., 2015; Mertens, Kuhn, Raes, Kalisch, De Houwer, & Lonsdorf, 2016; Mertens, Raes, et al., 2016; Raes, De Houwer, De Schryver, Brass, & Kalisch, 2014; Soeter & Kindt, 2012).

Furthermore, Field and colleagues (e.g., Field & Lawson, 2003; Field, Argyris, & Knowles, 2001) developed a Fear Beliefs Questionnaire to assess children's beliefs about animals that they had received verbal information about. This questionnaire consists of rating subjective feelings in situations involving the animals, such as whether the children would be happy to play with the animals or to be alone with them. Field and colleagues found that threatening verbal information (see Section 2.1.3), compared to neutral or positive verbal information, resulted in more fear beliefs about the animals as measured with this Fear Beliefs Questionnaire.

Finally, Muris and colleagues (e.g., Muris et al., 2009; Remmerswaal, Huijding, Bouwmeester, Brouwer, & Muris, 2014) investigated the effect of verbally threatening information on reasoning biases. Children were given threatening, positive, neutral or no information about an unknown animal. Afterwards, reasoning biases (e.g., confirmation biases, covariation biases) about this animal were assessed using contingency judgement tasks and by asking children what extra information about the animals they would like to hear. Muris and

colleagues found that children showed more reasoning biases for animals about which they had received threatening information.<sup>3</sup>

### **2.2.2. Physiological measures**

Common physiological responses that index conditioned fear are increased skin conductance responses (SCRs), potentiation of the startle reflex, heart rate acceleration, facial electromyography (EMG), changes in the event-related potential (ERP) components related to the processing of the CS and brain activity measured with fMRI (Lipp, 2006; Lonsdorf et al., 2017; Miskovic & Keil, 2012). Conditioned fear via verbal instructions has been observed using all these different psychophysiological measures: Increased SCRs (Cook & Harris, 1937; Costa et al., 2015; Mertens & De Houwer, 2016a), potentiation of the startle reflex (Costa et al., 2015; Mertens & De Houwer, 2016a), heart rate acceleration (Costa et al., 2015; Field & Schorah, 2007), facial EMG (Costa et al., 2015), changed ERP components (Bublitzky & Schupp, 2012; Weymar, Bradley, Hamm, & Lang, 2013) and neural activation in brain areas which are also typically activated for fear conditioning via CS-US pairings (Phelps et al., 2001). These studies demonstrate that verbal instructions can be a powerful manipulation to install conditioned physiological responses to CSs.

### **2.2.3. Behavioral measures**

Behavioral measures of conditioned fear include the physical distance to and degree of interaction with the CSs using behavioral approach tasks. Field and colleagues found that

---

<sup>3</sup> As we noted above, threatening information differs from conditioning instructions in that the former describes stimulus properties (e.g., has claws) rather than a spatio-temporal contingency (e.g., CS will be followed by US). We nevertheless refer to these studies in our review because threatening information can imply a CS-US contingency as well (e.g., animals with claws can inflict harm; see also: Muris & Field, 2010).



children took more time to touch a closed box in which the threatening animal supposedly was in and they put a toy figure representing themselves further away from the threatened animals in a toy animal park (Field & Lawson, 2003; Field & Storksen-Coulson, 2007). Similarly, Bennett et al. (2015) found that participants more often pressed the spacebar to remove the stimulus from the screen and avoid the US when a CS that was paired with threatening information was presented than when a control stimulus was presented (see also Dymond, Schlund, Roche, De Houwer, & Freegard, 2012).

Effects of fear conditioning via verbal instructions have also been investigated for behavioral indices of attentional bias. In a recent study, Deltomme, Mertens, Tibboel, and Braem (2017) found evidence for a visual attention bias as measured with the dot-probe task (MacLeod, Mathews, & Tata, 1986) for a CS conditioned via verbal instructions. That is, participants were faster to detect a dot on the side of the screen where the verbally conditioned CS was previously presented compared to detecting a dot on the screen where a control stimulus (i.e., a CS-) was previously presented. This result indicates that fear conditioning via verbal instructions can install visual attention biases.

Finally, effects of fear conditioning via verbal instructions have been observed on reaction time tasks that were developed to measure implicit preferences for stimuli. Specifically, in the study of Field and Lawson (2003), children were asked to classify animals about which they had received positive or threatening information together with positive and negative nouns and adjective in an Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; see De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009 for a review about this measure). The children displayed negative implicit evaluations for the animals they had received threatening

information about relative to the animals they had received positive information about.

## **2.3. Variations of the relationship between stimuli**

### **2.3.1. Extinction and return of fear**

Presenting CSs in the absence of the US typically results in a reduction of conditioned fear (i.e., extinction of fear). This procedure installs a change in the relationship between stimuli: A contingency between a CS and US is followed by the absence of a contingency between the CS and US, resulting in a decrease of CRs. Importantly, extinction has also been observed for fear conditioning via verbal instructions: When participants are first told that a CS will be followed by a US, subsequent presentations of the CS without the US reduces verbally conditioned fear responses (as measured with SCRs, US expectancy ratings, and fear ratings) to this CS (Mertens & De Houwer, 2016b; Mertens, Raes, et al., 2016). Furthermore, we found that extinction of verbally installed fear is sensitive to similar kinds of manipulations as fear installed through stimulus pairings. Specifically, a context switch (i.e., context renewal; Mertens & De Houwer, 2016b) and an unsignaled administration of the US (i.e., reinstatement; Mertens, Kuhn, et al., 2016; Mertens, Raes, et al., 2016) results in the return of fear after extinction, similar to what has been observed for fear conditioning via stimulus pairings (Haaker, Golkar, Hermans, & Lonsdorf, 2014; Vervliet, Baeyens, Van den Bergh, & Hermans, 2013).

### **2.3.2. Stimulus competition effects**

Another element of the CS-US relationship that impacts learning is the presence of a relationship of another CS with the same US. For instance, blocking of fear conditioning to a specific CS is observed when this CS is paired with the US in the presence of another CS that was previously paired with the US (Hinchy, Lovibond, & Ter-Horst, 1995; Kamin, 1969).

Lovibond (2003) demonstrated that such stimulus competition effects can also be obtained when participants are given written instructions about the pairings of the CSs and US instead of actually experiencing these pairings. Specifically, in a first phase, Lovibond (2003, Experiment 2) gave participants instructions about the relationship between compounds of two CSs and the presence of a shock (e.g., AB+ and CD+). In a second phase, participants received instructions about the relationship between a single CS and the absence of a shock (A-). This information presumably allowed participants to infer that CS B must have been the cause for the presence of the shock in the first phase (called 'the release from overshadowing' effect). Indeed, participants showed more fear (as measured by US expectancy ratings and SCRs) for CS B compared to CS D. The same results were obtained in another experiment in which the CSs and the CS-US pairings were actually presented to the participants (Lovibond, 2003; Experiment 1). Hence, stimulus competition effects can also be obtained when fear conditioning is established via verbal instructions.

## **2.4. Characteristics of the participants**

### **2.4.1. Trait anxiety**

Differences in trait anxiety appear to modulate the acquisition, extinction and return of conditioned fear (Chan & Lovibond, 1996; Grillon, Ameli, Foot, & Davis, 2002). Similar modulations of fear through trait anxiety have been observed for fear conditioning via verbal instructions. In a study by Grillon and colleagues (1993) participants were told that when one light (red or blue) was on they might receive an electric shock, and when another light was on (red or blue) they would not receive any shocks. They were also told that when the light that predicted a shock was on, they would only receive a shock in the last 10 seconds of a 50 seconds

period (during CS presentation a timer counted from 0 to 50 seconds). Grillon et al. (1993) observed that participants displayed more fear potentiated startle when the verbally conditioned light was on compared to the control light, and that this effect was stronger for high anxiety subjects. These findings thus indicate that fear conditioning via verbal instructions can be influenced by trait anxiety levels.

#### **2.4.2. Age of the participants**

Effects of verbal threatening information have not only been observed for university students in a laboratory setting, but also for school-aged children in more naturalistic settings. The above mentioned studies of Field (Field et al., 2001; Field & Lawson, 2003; see Section 2.1.3) demonstrated that threatening information about an animal increased fear beliefs, avoidance and negative implicit evaluation for the animal in 6 and 9 year old children. Hence, conditioned fear on the basis of verbal instructions can be established with young children (see Muris & Field, 2010, for a review of the effects of verbal threatening information in children).

#### **2.5. Elements of the broader context**

Environmental elements such as task demands (e.g., performing another attention demanding task) or distractors can interfere with the acquisition and extinction of conditioned fear established through CS-US pairings (e.g., Carter, Hofstotter, Tsuchiya, & Koch, 2003). To our knowledge, no studies have addressed the impact of such factors on fear conditioning via verbal instructions.

#### **2.6. Summary, Open Questions, and Conclusions**

Research demonstrating that conditioned fear responses can be installed by mere verbal instructions about the contingency between a CS and an aversive US was already reported 80

years ago (Cook & Harris, 1937). Since this early investigation, more effects of procedural variations on this type of fear conditioning have been discovered: it occurs with many types of CSs and USs, it is found with subjective, behavioral and physiological measures of fear, it is sensitive to the relationship between stimuli and it is modulated by the participants' characteristics. We will discuss the implications of these findings for theories of fear learning in Section 4.

However, despite a considerable amount of research, several open questions about the effects of procedural variations remain for fear conditioning via verbal instructions. Future research is needed to clarify whether verbally conditioned fear can be obtained with different types of (non-visual) CSs and (interoceptive) USs; whether effects can be found for specific types of outcome measures, such as perceptual discrimination, pupil dilation, and visual awareness; whether specific variations of the relationship between stimuli, such as latent inhibition, partial reinforcement, conditioned inhibition, trace conditioning, and forward and backward blocking impact fear conditioning via verbal instructions; what the role is of individual difference factors like intolerance of uncertainty, gender, hormone levels and specific genetic polymorphisms; and what the impact is of dual-tasks and distractors on fear conditioning via verbal instructions (see Lonsdorf et al., 2017, for a comprehensive overview of factors that moderate fear conditioning via CS-US pairings and that might also moderate fear conditioning via instructions). Furthermore, additional research should address the impact of perceptual generalization, stimulus masking, context switches, stimulus competition and trait anxiety on fear conditioning via verbal instructions, and further confirm the effects of fear conditioning via verbal instructions on CS valence ratings, reasoning biases, facial EMG, heart rate, neural activation, visual attention biases and implicit measures of evaluations (see Table 1). By highlighting these caveats in the literature,

we hope that our paper will inspire future studies on this important topic.

### **3. Moderating Effects of Verbal Instructions on Fear Conditioning via CS-US Pairings**

In any human fear conditioning procedure with CS-US pairings, participants are typically given instructions during different stages of the experiment. These instructions can have a tremendous impact on conditioning effects. In the next section, we give an overview of ways in which instructions about the elements of the fear conditioned procedure moderate the impact of CS-US pairings on conditioned fear. We will use the same five elements as before, focusing on the effects of verbal instructions about the stimuli, the outcome measures, the relationship between stimuli, the participants' characteristics and the elements of the broader context. An overview of the studies discussed in this section can be found in Table 2.

#### **3.1. Effects of instructions about the stimuli**

##### **3.1.1. Threatening instructions about the CSs**

At least three studies have investigated the effects of verbal threatening information about the CSs on fear conditioning via CS-US pairings. In the first study, Field and Storksen-Coulson (2007) gave children threatening information about an unknown animal (see Table 2 for a partial transcription of the threatening information). A subsequent negative conditioning experience with this animal (a sudden movement in a closed box that supposedly contained this animal) resulted in more fear (as measured with a behavioral avoidance task) compared to just this negative experience without the threatening information or just the information without the experience. In a second study by Ugland, Dyson and Field (2013; Experiment 1), students received threatening information about two animals and no information about two other animals. The subsequent pairing of one animal from each category with a loud aversive sound resulted in faster fear

conditioning (as measured by US expectancy ratings) for the previously threatened animal than for the neutral animal. Finally, in a study by Mertens and De Houwer (2017) participants received threatening information about two animals and control information about two other animals. Pictures of one threatening and one non-threatening animal were subsequently paired with an electrical shock according to a continuous (100%) or a partial (50%) reinforcement scheme (i.e., CS+s), and two other animals (one threatening, one non-threatening) were not paired with the shock (i.e., CS-s). This acquisition phase was then followed by an extinction phase. Mertens and De Houwer (2017) found that the threat instructions resulted in heightened fear for the threatening CS- compared to the non-threatening CS- (as measured by US expectancy ratings) and delayed extinction of the potentiated startle reflex for the threatening CS+, especially in the partial reinforcement condition. Together, these studies demonstrate that verbal threat instructions about the CSs can result in biased fear conditioning via CS-US pairings, as indicated by more strongly acquired fear (Field & Storksen-Coulson, 2007), faster acquired fear (Ugland et al., 2013) and delayed extinction of fear (Mertens & De Houwer, 2017).

### **3.1.2. Positive instructions about the CSs**

Not only threatening information, but also positive information appears to moderate conditioned fear established through CS-US pairings. In a study by Eifert (1984), the effects of positive self-instructions about the CS were investigated. Participants were first conditioned to pictures of snakes (i.e., the CSs) paired with an electrical shock. Thereafter, participants went through 20 extinction trials consisting of the pictures of the snakes without the electric shock and during which they additionally heard positive or negative statements about these snakes (e.g., “this pretty snake lies peacefully in the sun” or “this ugly snake is disgusting”), which they had to

repeat sub-vocally. Eifert (1984) found that positive verbalizations about the snakes decreased SCRs to the snakes (relative to SCRs to the snakes during a prior habituation phase) and resulted in more positive subjective evaluations of the snakes. In contrast, participants who made negative verbalizations about the snakes showed unaffected SCR and reported an increase of negative subjective evaluations of the snakes. Similar results were reported in a later study by Eifert and Schermelleh (1985).

More recently, Ugland et al. (2013, Experiment 2) conditioned participants to two unknown animals by repeatedly pairing pictures of these animals with an unpleasant sound. Two other animals were only paired with the sound for 50% of the trials. After this conditioning phase, participants received either positive information about two of these animals (a description of how approachable and friendly these animals are in a fake news bulletin) and no information about the two other animals. Ugland et al. (2013) found a decrease of conditioned fear beliefs as measured by the Fear Beliefs Questionnaire (see Section 2.1.2.) only for the animals about which they received positive instructions. Furthermore, an extinction phase after these positive instructions did not strengthen the effects of the positive instructions.

Finally, in a recent study by Luck and Lipp (2018), male faces were used in a differential fear conditioning procedure (i.e., one face was paired with an electrical shock and the other was not). Following the acquisition phase, participants received positive (e.g., “Ben has been raising money for a local homeless shelter”) or negative (e.g., “Chris is currently in jail for setting fire to his elderly neighbor’s house”) evaluative information about the faces. Luck and Lipp (2018) found that positive evaluative information about the CS+ immediately reduced negative conditioned evaluations and eliminated differential SCRs (though the latter effect was only found



in Experiment 1 and not Experiment 2, which may have been due to the negative evaluative information provided for the CS- in Experiment 1). Taken together, the studies of Luck and Lipp (2018), Ugland et al. (2013), and Eifert (Eifert, 1984; Eifert & Schermelleh, 1985) indicate that post-acquisition positive instructions about the CSs can counteract the effects of fear conditioning via CS-US pairings.

### **3.1.3. Generalization instructions**

Generalization of fear conditioned via CS-US pairings to perceptually similar stimuli can also be moderated by verbal instructions. For instance, in a study of Vervliet, Kindt, Vansteenwegen, and Hermans (2010), a yellow triangle and a shock were repeatedly paired. Crucially, before the conditioning phase, participants received instructions that the shape or the color of the stimulus was informative for the occurrence of the shock. In a subsequent test phase, participants' fear responses (as measured by US expectancy ratings and skin conductance responses) towards a blue triangle (same shape) and a yellow square (same color) were assessed. Generalization of fear to these test stimuli depended on the pre-acquisition instructions: More fear generalization to the blue triangle (relative to the yellow square) was found for the group that was instructed that the stimulus shape was informative for shock occurrence and more fear generalization to the yellow square (relative to the blue triangle) was found in the group that was instructed that the stimulus color was informative for shock occurrence. This finding was recently replicated by Ahmed and Lovibond (2015). However, rather than giving the generalization instructions before the learning phase, they provided these instructions after the learning phase, thereby demonstrating that the effects of Vervliet et al. (2010) were not merely the result of attention to a specific feature of the stimulus during the acquisition phase, but more likely reflect

an active reasoning process.

In another study by Boddez, Bennett, van Esch, and Beckers (2017), participants received instructions informing them that the likelihood of receiving an electric shock was lower when the image looked more similar to another image (of a black or white square) that was paired with the electric shock in the previous phase. This instruction resulted in an inversion of the typical generalization gradient in a subsequent test phase with generalization stimuli (squares in different shades of grey): Fear generalized less (as measured by US expectancy ratings) to stimuli that looked more similar to the CS+. However, CS valence ratings were not sensitive to the instructions and followed the typical generalization gradient regardless of the specific instructions.

Finally, a recent study by Scheveneels, Boddez, Bennett, and Hermans (2017) investigated the effect of instructions on the generalization of extinction to perceptually similar stimuli. Participants went through a conditioning phase with a particular CS (a 'Fribble'; i.e., an artificial animal-like figure; Barry, Griffith, De Rossi, & Hermans, 2014) followed by an extinction phase with a perceptually similar generalization stimulus (GS). Before a testing phase with yet another perceptual similar GS, participants were told that the GS during the extinction phase was a typical or an atypical exemplar of the Fribbles. The results demonstrated that the generalization instructions affected extinction to the GS in the test phase: lower US expectancy ratings were obtained for the GS in the test phase when the GS in the extinction phase was described as a typical exemplar of the Fribbles compared to when it was described as an atypical exemplar. The results from SCRs were in the same direction, but were not significant. Taken together, these results of Vervliet et al. (2010), Ahmed and Lovibond (2015), Boddez et al.

(2017) and Scheveneels et al. (2017) indicate that generalization of conditioned fear is affected by verbal instructions.

#### **3.1.4. Description of the US**

It is common in fear conditioning research to refer to the electric shock used in the experiments as “electrical stimulation”, “electrotactile stimulus” or “electrocutaneous stimulus”. Similarly, fear conditioning studies using unpleasant sounds also use variable terms such as “aversive sound” or “unpleasant sound”. Although different opinions may exist about the optimal description of the US, no study has formally investigated its effect on fear conditioning. On the one hand, the exact description of the US might be trivial because participants probably infer that the description refers to an electric shock or a loud noise, which, in most studies, they have been exposed to in a pre-experimental work-up procedure. On the other hand, it might be a non-trivial factor because it may influence whether potential participants decide to participate in the experiment (i.e., it could lead to a selection of non-anxious participants) and it might change the perception of the aversiveness of the US through a process of US inflation (Hosoba, Iwanaga, & Seiwa, 2001; White & Davey, 1989). Studies or a meta-analysis that formally investigate the effect of this factor are currently lacking.

#### **3.1.5. US reappraisal instructions**

Participants can also be given (positive) information about the US (e.g., Dibbets et al., 2012; Blechert et al., 2015). In a study by Dibbets et al. (2012), participants were asked to imagine that they had witnessed an accident of a child being hit by a car and that they were unable to call for help and rescue the child. Next, participants received pairings between a picture of a car (CS+) and a picture of a mutilated child (US) in a certain context (background picture of

a residential area). A picture of a motor bike (CS-) was not paired with the US. Then, some participants ('reappraisal group') were instructed to imagine that they called an ambulance after witnessing the accident and that the child was saved. Thereafter, participants received extinction trials in a new context (background picture of a service station) followed by a return to the acquisition phase in which CSs were presented again without the US. The most important result was that the reappraisal group showed a reduced return of conditioned fear after returning to the original acquisition context (as measured by US expectancy ratings) compared to a control group that was instructed to imagine an unrelated positive event. These results suggest that verbal instructions can allow for a change in the US representation and thereby reduce the return of conditioned fear after a context switch.

In two other relevant studies by Delgado, Nearing, LeDoux, and Phelps (2008) and Shurick et al. (2012), participants went through a differential conditioning phase in which pictures of spiders and snakes (Shurick et al., 2012) or colored shapes (Delgado et al., 2008) were paired with an electric shock, whereas another picture of a snake/spider or shape was not paired with the shock. In the study by Delgado et al., participants were sometimes asked to think about a calming nature scene while viewing the CS+ during the conditioning phase, while in the study by Shurick et al. (2012) half of the participants were told that they should not exclusively focus on the shock and focus on less negative aspects of the CS+ after the conditioning phase. These interventions, framed as reappraisal and cognitive restructuring, respectively, resulted in a decrease of conditioned SCRs to the CS+s compared to their control conditions (i.e., focusing on natural feelings in Delgado et al., 2008, and a card sorting task in Shurick et al., 2012). Thus, the results of the studies by Dibbets et al. (2012), Delgado et al. (2012) and Shurick et al. (2012) indicate that instructions aimed at reappraising the US can produce a reduction in conditioned

fear responses, can facilitate extinction and can reduce the return of conditioned fear responses after a context change.

### **3.1.6. US rehearsal instructions**

Finally, participants can be instructed to mentally imagine the US after CS-US pairings. Such instructions seems to delay extinction of conditioned fear (Dadds, Bovbjerg, Redd, & Cutmore, 1997). For instance, in studies by Davey and colleagues (Davey & Matchett, 1994; Jones & Davey, 1990), participants were instructed after a differential (CS+ vs CS-; pictures of a triangle or a kitchen tap, respectively) conditioning phase to mentally imagine the US from the previous phase whenever the word think was presented on the computer screen, after which they were presented with the CS+ image. Participants in two control groups were either asked to think of a cat meowing and their reactions to it as vividly as possible when they saw the word think (neutral control condition) or to think of someone trying to stick a pin into their eye (unrelated control condition). Davey and colleagues found that these US rehearsal instructions maintained differential conditioned SCRs in the experimental group compared to the control groups in a subsequent test phase. Related results were obtained in a study Drummond, White and Ashton (1978). These results of Davey and colleagues and Drummond et al. (1978) indicate that post-acquisition US rehearsal instructions can help to maintain conditioned fear responses.

Table 2. Overview of the available published articles investigating the moderating effects of verbal instructions on fear conditioning via CS-US pairings.

Procedural elements		References	Example of the instructions
<b>Instructions about the stimuli</b>	Threatening instructions about the CS	Field & Storksen-Coulson (2007)* Ugland et al. (2013) Mertens & De Houwer (2017)	<i>Quolls are very dangerous, and live in dark places in the woods, where they hunt other creatures with their long sharp teeth and claws.</i>
	Positive instructions about the CS	Eifert (1984) Eifert & Schermelleh (1985) Ugland et al. (2013)* Luck & Lipp (2018)	<i>We have been introduced to two of the friendliest animals that you are ever likely to come across in the wild! [...] They have become increasingly popular as pets as they are quite easy to look after and they are great with children.</i>
	Generalization instructions	Vervliet et al. (2010)* Ahmed & Lovibond (2015) Boddez et al. (2017) Scheveneels et al. (2017)	<i>The COLOUR of the geometrical figures is important to know when the electrical stimulus will follow.</i>
	Description of the US	Studies are available, but have not been compared systematically.	<i>Shock, electrical shock, electro-tactile stimulus, electrocutaneous stimulation, ...</i>
	US reappraisal instructions	Delgado et al. (2008) Dibbets et al. (2012)* Shurick et al. (2012)	<i>After a car picture we want you to close your eyes and imagine the complete scene in which you save the boy.</i>
	US rehearsal instructions	Drummond et al. (1978) Jones & Davey (1990)* Davey & Matchett (1994)	<i>Whenever the word 'think' is presented, think about the loud tone presented during phase I.</i>
<b>Instructions about the outcome measures</b>	Instructions to regulate emotional reactions	Hill (1967) Dawson & Reardon (1969) Swenson & Hill (1970) Harvey & Wickens (1971) Lissek et al. (2007)*	<i>Whatever fear you might experience in response to the picture, if you are instructed to ENHANCE, we would like you to increase the intensity of fear you feel.</i>
<b>Instructions about the regularity</b>	Contingency instructions	Dawson et al. (1979) Tabbert et al. (2011) Javanbakht et al. (2016) Duits et al. (2017)*	<i>Shocks will only be administered during presentation of the picture presented above.</i>
	Instructed extinction	Bridger & Mandel (1964; 1965)	<i>In the next phase you will</i>

		Hugdahl & Öhman (1977)* Sevenster et al. (2012) Luck & Lipp (2015a; 2015b)	<i>not receive any more shocks.</i>
	Contingency reversal instructions	Wilson (1968) Grings et al. (1973) McNally (1981)* Mertens & De Houwer (2016a) Atlas et al. (2016)	<i>By now you will have noticed that the shock occurs only after the snake slide. From now on, the procedure is reversed; the shock will occur only after the spider slide.</i>
	Stimulus competition instructions	Lovibond (2003) Boddez et al. (2013)*	<i>X is a stimulus causing shock and X was causing a shock during training.</i>
	Instructions supporting inferential reasoning	Raes et al. (2011)* Zeng et al. (2015)	<i>There was a technical disturbance with the task. Through this disturbance, no noise had been presented for some time. However, the disturbance is now solved and the experiment will now continue.</i>
	CS-US rehearsal instructions	Yaremko & Werner (1974) Joos et al. (2012a; 2012b)*	<i>Think back to the picture, the scream and the relationship between them.</i>
<b>Instructions about the characteristics of the participants</b>	Instructions about coping abilities	Zlomuzica et al. (2015)*	<i>According to our analyses, you are in the top 1% of “copers”. In general you have fewer negative emotions and recover much more quickly, and you feel capable of overcoming difficult life events in the future.</i>
<b>Instructions about elements of the broader context</b>	Instructions about context cues	Neumann (2007)* Mertens & De Houwer (2016b)	<i>You may also notice that the colours of the lights and the background sound in the room may also change. It is very important to note that the changes in a signal's meaning and the changes in the lights and sounds of the room are NOT RELATED.</i>

---

Note: An asterisk refers to the study we used the instructions from as an example. Further note that verbatim instructions were sometimes not available. In this case we used the description in the procedure to paraphrase the instructions.

### **3.2. Instructions about the outcome measures**

Participants can be instructed to regulate their fear responses conditioned via CS-US pairings. For instance, Hill (1967) told half of the participants to “prepare physiologically” for the possibility of a shock and that “the intelligent thing to do is to become conditioned” (facilitatory instructions). The other half of the participants were told that “it really doesn’t make much sense to continuously respond to the tone” and that “the intelligent thing to do is to not become conditioned” (inhibitory instructions). In a single-cue fear conditioning procedure (i.e., only a CS+), Hill observed that the facilitatory instructions group consistently demonstrated stronger conditioned SCRs. These observations of Hill were confirmed in several follow-up studies (Dawson & Reardon, 1969; Harvey & Wickens, 1971; Swenson & Hill, 1970).

In a more recent study, Lissek et al. (2007) investigated the emotional regulation of the startle response in a threat-of-shock paradigm (which included occasional shock administrations, hence acquired fear was partly based on CS-US pairings). Participants were told to either “suppress”, “maintain” or “enhance” their emotional responses while viewing threatening and safety cues (the words ‘shock’ or ‘safe’). Furthermore, participants were promised \$50 if they successfully managed to regulate their emotions. Participants were successful to either suppress or enhance their startle responses compared to the maintain condition. These results of Lissek et al. (2007) thus further demonstrate that verbal instructions can successfully moderate physiological responses conditioned via CS-US pairings.

### **3.3. Instructions about the relationship between stimuli**

#### **3.3.1. Contingency instructions**

Fear conditioning studies vary in whether they give participants information about the



contingencies in the experiment. Giving participants such instructions about the presence of contingencies (either specifying them exactly, or informing participants that there is a contingency without specifying) is generally believed to facilitate the acquisition of conditioned fear responses (Colgan, 1970; Grings & Kimmel, 1959; Kimmel & Pennypacker, 1963; Norrholm et al., 2006; Warren et al., 2014; Duits et al., 2017). Indeed, several studies have demonstrated that either giving participants contingency instructions prior to acquisition (Atlas, Doll, Li, Daw, & Phelps, 2016; Dawson, Catania, Schell, & Grings, 1979; Javanbakht et al., 2016; Tabbert et al., 2011) or during acquisition (Duits et al., 2017) facilitates conditioned fear acquisition via CS-US pairings on both subjective and psychophysiological measures of fear relative to a condition where no contingency instructions are provided. Unfortunately, so far no meta-analysis has been conducted to systematically compare studies in which fear conditioning via CS-US pairings was conducted either with or without contingency instructions (for a meta-analysis of the effects of contingency instructions on brain activation in fear conditioning via CS-US pairings, see Mechias et al., 2010).

### **3.3.2. Instructed extinction**

Similar to how participants can be instructed about the presence of certain contingencies in an acquisition phase, participants can be instructed about the absence of a contingency prior to an extinction phase (Bridger & Mandel, 1965; Hugdahl & Öhman, 1977; Swenson & Hill, 1970). Overall, most of these studies demonstrate that these instructions immediately and completely abolish conditioned fear on all measures. However, some studies found that (some) conditioned fear responses can persist after these instructions (Bridger & Mandel, 1964, 1965; Sevenster, Beckers, & Kindt, 2012). One crucial variable here might be whether or not electrodes for shock

delivery are removed or not. When extinction instructions are combined with the removal of the shock delivery electrodes, these instructions usually result in the reduction of conditioned fear responses (Luck & Lipp, 2016a; though see Luck & Lipp, 2015b, for a study demonstrating that removal of the shock electrodes did not moderate the instructed extinction effect). However, there might still be some limitations to the effects of instructed extinction. Specifically, other crucial factors that seem to determine whether instructed extinction is complete are whether evaluative judgments are used as the CR, whether a highly intense US is used and whether evolutionary fear-relevant CSs (i.e., pictures of snakes and spiders) are used in the conditioning procedure. Under these conditions, instructed extinction seems to be less complete (e.g., Hugdahl & Öhman, 1977; for a review of instructed extinction studies see Luck & Lipp, 2016a).

### **3.3.3. Contingency reversal instructions**

Another situation where participants can be informed about changed contingencies is in a contingency reversal situation (Atlas et al., 2016; Grings, Schell, & Carey, 1973; McNally, 1981; Mertens & De Houwer, 2016a; Wilson, 1968). Here, participants are informed after a differential conditioning phase that the contingencies from the previous phase will be reversed in the subsequent phase. Research using this procedure has demonstrated that these instructions successfully reversed conditioned SCRs with neutral CSs (Grings et al., 1973; Wilson, 1968) and fear-relevant CSs (i.e., pictures of snakes and spiders) (Atlas & Phelps, 2018; McNally, 1981). Mertens and De Houwer (2016a) have recently extended these studies by demonstrating that also conditioned potentiated startle reflexes can be reversed with contingency reversal instructions. Moreover, a recent study by Luck and Lipp (2016b) indicates that also conditioned evaluative responses can be influenced by contingency reversal instructions (though this effect seems less

outspoken and may require certain preconditions such as habituation trials of the CS+).

#### **3.3.4. Stimulus competition instructions**

As mentioned before (see Section 2.3.2), the pairing of one CS with a certain US can moderate the effectiveness of another CS to become conditioned to this US (Kamin, 1969). Also instructions about the pairing between a CS and the US can moderate fear conditioning via CS-US pairings for another CS. For instance, Lovibond (2003; Experiment 3) gave participants verbal instructions telling them that “A is safe” after a conditioning phase in which compounds of CSs (AB+ and CD+) were paired with an electric shock. This simple instruction about CS A resulted in more fear (as measured by SCRs and US expectancy ratings) for CS B compared to CS D in a subsequent test phase, most likely because it allowed participants to infer that CS B must have been the stimulus causing the shock in the first part of the experiment (see Section 2.3.2).

In a related study of Boddez, Baeyens, Hermans, Van der Oord and Beckers (2013) participants also received pairings between CS compounds (AB+ and CD+) and an electric shock. Thereafter, participants received instructions that “A is a stimulus causing shock and A was causing a shock during training”. These instructions resulted in *less* fear (as measured by US expectancy ratings) for CS B compared to CS D. Unfortunately, fear conditioning was not found for SCRs during the acquisition phase in the study of Boddez et al. (2013) and therefore the effects of the instruction manipulation on this measure could not be interpreted. Nevertheless, the combined studies of Lovibond (2003) and Boddez et al. (2013) provide persuasive evidence that information about the relationship between one CS and the US can moderate fear conditioning via CS-US pairings.

### **3.3.5. Instructions supporting inferential reasoning**

The studies on stimulus competition already indicate that reasoning processes may be involved in fear conditioning via CS-US pairings (Lovibond, 2004; Lovibond, Mitchell, Minard, Brady, & Menzies, 2009). Specifically, participants appear to use inferential reasoning to determine which CS is a good and logically plausible predictor of the US. Another demonstration of this idea was provided by Raes, De Houwer, Verschuere and De Raedt (2011). In their study, participants were first conditioned by pairing a CS (a colored square) repeatedly with a loud noise, while another CS (another colored square) was not paired with the loud noise. Thereafter, participants went through an extinction phase during which the loud noise was no longer presented. Importantly, after several extinction trials the experiment was interrupted and participants were told that there had been a technical error. Half of the participants were told after one minute that the error had been fixed, after which they continued to the test phase, whereas the other half of the participants were told that the error had been fixed and that this error had resulted in a disruption of the administration of the sounds (the US), after which they continued to the test phase. Raes et al. (2011) found that the participants who were told that the error disrupted the administration of the US showed a strong increase in differential fear responses in the test phase as measured by SCRs and US expectancy ratings, compared to the control group. The results of Raes et al. (2011) thus demonstrate that instructions that allow for inferences about the absence of the US during an extinction phase can allow for an immediate return of fear. These results were recently independently replicated (Zeng, Jia, Wang, Zhang, Liu, & Zheng, 2015).

### **3.3.6. CS-US rehearsal instructions**

Participants can also be instructed to mentally rehearse the CS-US pairings (Dadds et al.,

1997; Joos, Vansteenwegen, & Hermans, 2012a; 2012b). In a study by Yaremko and Werner (1974) one group of participants were exposed to one tone (CS)-shock (US) pairing, whereas another group of participants (a pseudo-conditioning group) were exposed to the tone and the shock separated in time. After this exposure to the tone and shock, half of the participants in each group were told to imagine the tone being immediately followed by the shock when the experimenter said “tone shock”. The other half of the participants in each group were instructed to imagine the tone and shock separated in time when cued by the experimenter. Yaremko and Werner (1974) found that when the participants were exposed to the tone, the participants in the tone-shock pairing group showed stronger SCRs to the tone. More importantly for the present purposes, the group that was told to imagine the tone-shock pairings showed stronger SCRs to the tone compared to the group that was told to imagine the tone and shock separated in time. Finally, the group that was exposed to the tone and shock separated in time, but that was instructed to imagine the tone-shock paired in time, also showed stronger SCRs to the tone compared to its control group. Similar results of post-acquisition CS-US rehearsal instructions were reported by Joos et al. (2012a; 2012b) with fear and US expectancy ratings as outcome measures. These results of Yaremko and Werner (1974) and Joos et al. (2012a; 2012b) thus indicate that CS-US rehearsal instructions can strengthen the effects of CS-US pairings.

### **3.4. Instructions about the characteristics of the participants**

One demonstration of the effects of instructions about the characteristics of the participants on conditioned fear established through CS-US pairings is the study of Zlomuzica, Preusser, Schneider, and Margraf (2015). In their study, Zlomuzica et al. (2015) provided participants with fake feedback about their coping abilities after a differential fear conditioning

phase and prior to a fear extinction phase. Specifically, half of the participants were told that they were amongst the top one percent of copers and that they had excellent abilities dealing with stressful situations. The other half of the participants in the control group did not receive these verbal instructions. Zlomuzica et al. (2015) observed that these instructions facilitated extinction of conditioned fear as measured by SCRs and stimulus valence ratings compared to the control group. Interestingly, debriefing the participants about the fact that they had received fake information about their coping potential did not undo the effects of the prior instructions in a subsequent test phase. Hence, this study by Zlomuzica et al. (2015) illustrates that verbal instructions about the characteristics of the participants (coping potential) can have an impact on the extinction of conditioned fear established via CS-US pairings.

### **3.5. Instructions about the broader context**

As mentioned earlier (see Section 2.5), distractors and task demands can interfere with the acquisition of conditioned fear via CS-US pairings. To our knowledge, no study has been conducted in which participants are instructed about these elements of the procedure. However, two studies have looked at the effects of instructions about context cues on the context renewal effect (see Section 2.3.1) (Mertens & De Houwer, 2016a, Neumann, 2007). The experiments by Neumann (2007) indicate that verbal instructions that informed participants that context cues were irrelevant did not attenuate the context renewal effect. On the other hand, the experiment by Mertens and De Houwer (2016b) showed that instructions informing participants that context cues are relevant strengthened the context renewal effect. These two studies suggest that instructions about context cues may strengthen, but not attenuate, the context renewal effect. However, because these two studies did not use a fear conditioning procedure via CS-US pairings

(but instead used a contingency learning task in a video game and fear conditioning via verbal instructions, respectively) we will not discuss these studies further, but merely mention them here for completeness.

### **3.6. Summary, Open Questions, and Conclusions**

In the previous section we reviewed the research about the moderating effects of verbal instructions on fear conditioning via CS-US pairings. As this review illustrates, verbal instructions can have a tremendous impact. That is, verbal instructions are able to strengthen or reduce the acquisition of conditioned fear, can facilitate or delay the extinction of conditioned fear and can either facilitate or reduce the return of extinguished fear. Furthermore, these moderating effects of verbal instructions have been obtained for a wide variety of types of conditioned fear responses, a wide variety of conditioning procedures and have been found in different labs and by different researchers, which speaks for the robustness of the effects of verbal instructions on fear conditioning via CS-US pairings.

Despite this wealth of empirical work available, we have also noted a number of research topics which have not yet been investigated or that require more research. Specifically, more work is required to determine what the optimal description of the US is, what the meta-analytic effects of contingency instructions are, whether information about characteristics of the participants can moderate conditioned fear, and whether instructions about task demands, context cues, and distractors can moderate fear conditioning via CS-US pairings.

Finally, we should note that it is unlikely that we were able to survey all of the relevant research. Any conditioning experiment with humans typically involves giving participants instructions about the different elements in the procedure. It is probably the case that many

researchers piloted the effects of instructions more or less formally to determine which instructions are optimal for their purposes. However, most of these pilot studies remain unpublished and therefore unavailable to the scientific community. Nevertheless, our review should allow readers to form a good impression of the wealth of evidence that is available about the impact of verbal instructions on fear conditioning via CS-US pairings. Furthermore, the procedural approach we used here could guide researchers to new ideas and unexplored research questions.

#### **4. Integration with Mental Process Models of Fear Learning**

Mental process models of fear learning differ with regard to their emphasis on automatic versus controlled processes that are involved in fear learning. There are two basic models: dual-process (a.k.a. dual-route) models and single process expectancy models. According to the dual-process model, fear learning takes place, for a large part, through relatively uncontrolled and reflex-like fear learning processes in the amygdala (Grillon, 2009; LeDoux, 2014; Öhman & Mineka, 2001; Olsson & Phelps, 2007). Researchers have proposed that this subcortical automatic fear learning system operates independently of higher-order cognitions and language (Mineka & Öhman, 2002; Öhman & Mineka, 2001) and mediates learning via CS-US pairings and, to a lesser extent, learning via social observation (Grillon, 2009; Olsson & Phelps, 2007). Fear learning can also take place through controlled processes, but these mainly produce subjective beliefs and feelings of fear. Intense physiological and behavioral fear responses (defensive reflexes) are thought to result primarily from a low-level, fast, and efficient learning system that operates on CS-US pairings without requiring awareness, goals or time to process these pairings. Although such strong versions of dual-process models have been put forward in



the literature (e.g., LeDoux, 2014), there are also weaker versions that do allow an impact of verbal instructions and controlled processes on some physiological reactions (such as SCRs; e.g., Hamm & Weike, 2005). In contrast to these dual-process models, single-process expectancy models propose that learned fear is the result of conscious expectations of an aversive event in the presence of CSs (Davey, 1992; Lovibond, 2011; Mitchell et al., 2009; Reiss, 1980) which are formed in a controlled, effortful, slow and conscious way (Lovibond, 2011; Mitchell et al., 2009)<sup>4</sup>. Only when participants (consciously) learn to expect an aversive US in the presence of the CS will they show conditioned fear to the CS, in all fear response systems alike.

These two theories of fear learning differ considerably in the extent to which they consider verbal instructions to be involved in fear learning. Contrasting these two models of fear learning<sup>5</sup> allows us to derive testable hypotheses about the effects of verbal instructions on fear conditioning that can be evaluated with the research findings we described above. First, according to the dual-process models, verbal instructions should be unable or less able to install physiological and behavioral conditioned fear responses compared to fear learning via CS-US pairings (Grillon, 2009; LeDoux, 2014; Olsson & Phelps, 2007). In contrast, according to single-

---

<sup>4</sup> Note that we only consider expectancy models here in which expectancies are learned in a controlled, effortful and conscious way. Some models argue that expectancies can be *consequences* of conditioned fear responses (i.e., through the interpretations of these conditioned responses) which were installed through an automatic fear learning process, rather than that expectancies are the *cause* of these conditioned fear responses (e.g., Mineka & Ohman, 2002). These latter models are in fact dual-process models (for an extensive discussion of these two basic models see Lovibond, 2011). Also note that although expectancies are generally assumed to play a crucial role in fear learning, some have argued that conscious learning about CS-US contingencies might also influence behavior in the absence of conscious expectancies (e.g., Mitchell et al., 2009).

<sup>5</sup> We acknowledge that there is considerable variance between the different models within the classes we defined here, such as, for instance, the exact learning rule that the automatic learning process uses and the relative involvement of heuristics and logical rules in the controlled learning process. Nonetheless, a major difference between these models remains the relative contribution of automatic versus controlled processes and the role of language and instructions (McLaren et al., 2014; Mitchell et al., 2009; Shanks, 2010).

process expectancy models of fear learning, verbal instructions should, in principle<sup>6</sup>, be as effective to install fear responses in any of the fear response systems (see Section 2.2) as CS-US pairings (Lovibond, 2003, 2011). Second, if both fear learning through verbal instructions and through CS-US pairings are mediated by the same mental process as proposed by single-process models, fear learning through these two pathways should show similar procedural properties (i.e., they should be moderated in a similar way by elements of the procedure; Lovibond, 2003). In contrast, if fear learning through verbal instructions and through CS-US pairings are mediated by different mental mechanisms, these two pathways should often be differently affected by procedural variations (e.g., McLaren et al., 2014). Finally, a third hypothesis is that according to dual-process models, verbal instructions should be unable to interrupt or alter defensive fear responses installed via CS-US pairings because these two pathways are mediated by processes that operate independently from each other (Mineka & Öhman, 2002). In contrast, according to single-process expectancy models, information from different learning pathways are continuously integrated into one learning process and therefore should strongly interact (Lovibond, 2003). These three hypotheses will be evaluated here.

First, with regard to the effects of instructions in the absence of CS-US pairings, many studies with a wide variety of dependent variables (see Section 2.2 and Table 1) refute the idea that defensive conditioned fear responses can be installed only through CS-US pairings. Even for measures of conditioned fear that have been regarded as highly automatic defensive reflexes,

---

<sup>6</sup> For verbal instructions to have strong effects a number of preconditions may have to be fulfilled, such as that they are provided by a credible source, are logically plausible and are personally relevant (e.g., De Houwer, Hughes, & Brass, 2017). These preconditions probably do not apply for learning via CS-US pairings because these pairings, if experienced by the participants, are usually already credible, logical and personally relevant. However, the fact that learning through these two different pathways may be subjected to different preconditions to have effects does not need to implicate that information from these two different pathways is processed by different mental processes.

such as the startle reflex (Hamm & Weike, 2005), strong moderating effects of verbal instructions have been observed (Costa et al., 2015; Grillon, Ameli, Woods, Merikangas, & Davis, 1991; Mertens & De Houwer, 2016a). Hence, CS-US pairings may not be required to install defensive fear responses.

Second, our review reveals a striking similarity between the procedural properties of fear conditioning via verbal instructions and fear conditioning via CS-US pairings, which also speaks in favor of single-process expectancy models of fear learning. These properties include effects of stimulus preparedness, stimulus generalization, extinction, stimulus competition, context renewal, reinstatement, and effects of trait anxiety. Indeed, of all the different procedural elements we have considered (see Table 1), there was only one for which initial evidence revealed a difference between fear learning via verbal instructions and fear learning via CS-US pairings: the effects of stimulus masking on the expression of conditioned fear (Olsson & Phelps, 2004). Such an extensive overlap regarding the moderating effects of elements in the environment on learning suggests that the same mental processes mediate these two learning pathways (Lovibond, 2003).

Finally, many studies have demonstrated that verbal instructions can have a tremendous moderating impact on conditioned fear established through CS-US pairings. Verbal instructions can enhance or reduce fear conditioning via CS-US pairings, both when instructions are given prior (Field & Storksen-Coulson, 2007) or after the learning phase (Ugland et al., 2013). Furthermore, verbal instructions can impact the retention (Dibbets et al., 2012), generalization (Vervliet et al., 2010) and expression (Raes et al., 2011) of conditioned fear acquired via CS-US pairings. Again, these observations are not consistent with the strong versions of dual-process

models of fear conditioning. That is, dual-process models have difficulties with accounting for interactions between verbal instructions and defensive fear responses established through CS-US pairings because they presume that the processing of information from these different pathways is performed largely independently (Lovibond, 2003). In contrast, such findings are expected by single-process expectancy models of fear learning because information from different pathways is continuously integrated to create expectancies about the occurrence of aversive events.

However, a number of caveats of this theoretical analysis should be highlighted. First, in line with recent proposals (e.g., LeDoux & Pine, 2016) and for the sake of clarity, we focused on strong versions of dual-process models that make a strict distinction between outcome measures that are proposed to be primarily affected by low-level learning processes (i.e., defensive reflexes) and outcome measures that are proposed to be mainly affected by expectancies (e.g., US expectancy ratings, subjective fear ratings). However, according to several weaker versions of dual-process models (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Hamm & Weike, 2005), certain measures (such as SCRs and evaluative ratings) may be affected by both expectations and low-level learning processes. Interactions between fear conditioning via CS-US pairings and verbal instructions can be consistent with such weaker versions of the dual-process model because instructions may moderate the part of the conditioned fear response that is installed via expectancies. It is important to note that single-process expectancy models can also be consistent with dissociations between outcome measures and a lack of effects of verbal instructions because different outcome measures may be affected differently by different kinds of expectancies. For instance, the startle response may be affected more by quick judgements of the probability of an aversive event in a very specific situation, whereas evaluative ratings are affected more by judgements about the relationship between CSs and aversive events over a very

long time interval. The challenge for both classes of models is thus to specify more exactly under which conditions (and for which measures) interactions between fear conditioning via CS-US pairings and verbal instructions are expected (for a more extensive discussion of this point, see Lovibond, 2003, 2011).

A second caveat, specifically concerning the second hypothesis, is that many similarities between fear learning through instructions and through CS-US pairings are based on comparisons between different experiments, rather than within the same experiment. Only a few studies have directly compared the different learning pathways (Braem et al., 2017; Lovibond, 2003; Mertens, Kuhn, et al., 2016; Mertens & De Houwer, 2016a, Olsson & Phelps, 2004; Raes et al., 2014). This may lead to incorrect conclusions about similarities and differences between these two pathways due to uncontrolled differences between the studies. Hence, more studies are needed that directly compare the two learning pathways.

Finally, a third caveat is that a number of studies that investigated the moderating effects of verbal instructions on fear learning via CS-US pairings found support for dual-process theories of fear learning (e.g., Atlas et al., 2016; Bridger & Mandel, 1964; Hugdahl & Öhman, 1977; Sevenster et al., 2012). For instance, Hugdahl and Öhman (1977) observed that instructed extinction did not attenuate conditioned fear to fear-relevant CSs (i.e., pictures of snakes and spiders), even though the same instructions completely abolished conditioned fear to fear-irrelevant CSs. This result seems to indicate that part of the CR, especially to fear-relevant CSs, is determined by a learned representation that is unaffected by verbal instructions, in line with the predictions of a dual-process model of fear learning. Likewise, Sevenster et al. (2012) found that differential startle responses established through CS-US pairings were not immediately

eliminated through instructed extinction (whereas conditioned SCRs were immediately reduced). This result also support the prediction of dual-process models of fear learning that defensive responses (such as the startle reflex) are less sensitive to verbal instructions. Finally, several studies have found that verbal instructions do not influence conditioned negative valence ratings while they do impact other fear responses (Boddez et al., 2017; Luck & Lipp, 2015a; 2015b; see Sections 3.3.3 and 3.3.2). These findings may also be taken as indicative of a dual-process model, in which low-level associations primarily determine conditioned evaluative responses while conscious expectancies primarily determine the other fear responses (see Baeyens et al., 1992; though see Boddez et al., 2017, for a different interpretation of this dissociation). Nevertheless, the results of these studies are not uncontested. That is, other studies have demonstrated that verbal instructions can moderate conditioned fear installed via CS-US pairings to fear-relevant CSs as well (Atlas & Phelps, 2018; McNally, 1981; McNally & Foa, 1986; for a review of these studies see McNally, 1987). Likewise, several studies have demonstrated that verbal instructions can abolish differential startle responses established through CS-US pairings (Luck & Lipp, 2015a; 2015b; Mertens & De Houwer, 2016a). Finally, two recent studies have indicate that verbal instructions may in fact change conditioned negative valence ratings (Luck & Lipp, 2016b, 2018), though more research is required to determine under which conditions conditioned evaluations are sensitive to verbal instructions.

To conclude, research on the effects of verbal instructions on fear learning largely corresponds with single-process expectancy models of fear conditioning because (1) verbal instructions are highly effective to install subjective, physiological, and behavioral fear responses in the absence of any CS-US pairings, (2) fear learning via verbal instructions and via CS-US pairings have similar procedural properties (i.e., are similarly moderated by the elements of the

procedure, see Table 1), and (3) verbal instructions and fear learning through CS-US pairings strongly interact. Nonetheless, more theoretical development is required to specify exactly under which conditions effects of verbal instructions are expected, more research is needed which directly compares fear conditioning via verbal instructions and via CS-US pairings to establish possible systematic differences, and several findings providing initial, though contested, support for dual-process models require further research.

### **5. Limitations, Open Questions and Future Directions**

Being a narrative review, our paper does not provide a quantitative meta-analysis of the effects of verbal instructions on fear conditioning. Although it might be interesting to conduct meta-analyses of subsections of the literature that we reviewed, the range of phenomena that we covered in our review was too broad to allow for a comprehensive meta-analysis. Instead, we hope that our paper provides a useful overview of what we currently do and do not know about the role of instructions in fear learning. Another advantage of our approach is that it allowed a systematical comparison of fear learning via verbal instructions to fear learning via CS-US pairings.

A second limitation of our review is that we did not place much emphasis on the neural implementation of fear conditioning via verbal instructions. Several studies have addressed the neural correlates of fear conditioning via verbal instructions (see Mechias et al., 2010 for an overview and a meta-analysis of relevant studies), but we mostly limited our review to studies that have focused on subjective, behavioral and psychophysiological measures of conditioned fear. Interested readers are referred to Olsson and Phelps (2007) for a review of relevant studies and a neural model, and Mechias et al. (2010) for a relevant meta-analysis.

A third limitation of the research we have reviewed here is that explicit instructions may install experimental demand effects (i.e., participants may change their behavior in line with what they perceive to be the expectations of the experimenter based on the explicit instructions; Orne, 1962). We see two major arguments against such an interpretation of the research presented here. First, demand effects can arise only if participants are aware of the hypothesis, that is, if they know what pattern of results the experimenter wants to see. Although participants may have knowledge of basic effects (e.g., conditioning, extinction), it is unlikely that they are familiar with some of the more complex effects, such as biological preparedness effects, stimulus generalization, return of fear manipulations (renewal and reinstatement), stimulus competition, and effects of trait anxiety. Hence, it seems unlikely that, for these phenomena, participants know how they are expected to behave, which is a prerequisite for experimental demand. Second, all of the studies that we reviewed included either behavioral or psychophysiological measures of fear, which are less likely to be influenced by experimental demand. Hence, it is unlikely that experimental demand explains all the effects of verbal instructions in fear conditioning, although additional studies are required to directly test this possibility.

With regard to future research, open questions about the effects of verbal instructions in fear conditioning were indicated throughout this manuscript. For an overview of these open questions with regard to fear conditioning via verbal instructions and the moderating effects of verbal instructions on fear conditioning via CS-US pairings we refer readers to Sections 2.6 and 3.6, respectively. Furthermore, as we indicated in the previous section, more research is needed that directly compares fear conditioning via verbal instructions and via CS-US pairings to be able to assess any qualitative and quantitative differences between fear learning through these two pathways. Most of the similarities and differences between the two learning pathways that were



discussed in this review were based on comparisons between studies rather than within the same study. This, of course, entails a high risk of drawing incorrect conclusions about differences and similarities between the two learning pathways because of uncontrolled differences between studies. More recent studies have provided useful procedures to allow for such a comparison (Atlas et al., 2016; Braem et al., 2017; Mertens, Kuhn, et al., 2016; Raes et al., 2014).

Another important issue that was not directly addressed in this review is the impact of what participants are asked to do with the instructions that they receive. For instance, in the studies about imagery, participants were encouraged to actively imagine the instructions (e.g., Dibbets et al., 2012; Jones & Davey, 1990; Joos et al., 2012). However, such instructions do not seem to be necessary for verbal instructions to have an effect on conditioned fear (e.g., Mertens & De Houwer, 2016a; Vervliet et al., 2010). A related issue concerns the effects of believability of the instructions. Participants can be asked about the extent to which they believed the instructions (e.g., Raes et al., 2014), but none of the studies reviewed above have systematically investigated the impact of this factor on their results. Future studies investigating the effects of these two factors (imagery and believability) might determine possible limitations of the effects of instructions on conditioned fear. In fact, it might be interesting to relate research on fear learning via instructions to other types of learning via instructions. For instance, many of the variables that are known to be relevant in persuasion research (i.e., attitude learning via verbal instructions) might also be important for fear conditioning via verbal instructions (see De Houwer, Hughes, & Brass, 2017, for a framework that relates different areas of research on learning via instructions).

One important future direction of research about the effects of verbal instructions on fear

learning is to elucidate how verbal instructions give rise to maladaptive fears. There is strong evidence indicating that verbal information contributes to the development of pathological fear (Field, 2006; King, Eleonora, & Ollendick, 1998; Schindler, Vriends, Margraf, & Stieglitz, 2016). Hence, verbal instructions can be considered to be an important etiological pathway for pathological fear (Rachman, 1977, 1991), but it remains a relatively understudied pathway. More laboratory studies that focus on how fear can be established through verbal instructions, particularly with participants who are at a greater risk to develop anxiety disorder (such as people with elevated state and trait anxiety or elevated intolerance of uncertainty; e.g., Morris, Christakou, & van Reekum, 2015), can produce important insight into the conditions and the mechanisms that allow for the development of pathological fear (see also Muris & Field, 2010).

Finally, another important future direction is to further investigate how verbal instructions may be used to reduce conditioned fear and relate these results to how maladaptive or pathological fear can be remedied. Specifically, instructions in the lab can resemble interventions that are used in therapeutic settings to treat pathological fear (e.g., Blechert et al., 2015; Dibbets et al., 2012; Luck & Lipp, 2016a; Duits et al., 2017). Thus, there are opportunities to further investigate whether verbal techniques that are thought to be effective in therapeutic settings (e.g., reappraisal, cognitive restructuring) in a laboratory setting to validate these techniques and to determine optimal conditions for their effectiveness. Furthermore, besides determining optimal conditions, studies focusing on the role of verbal instructions in laboratory fear conditioning could also help to gain a better insight into the working mechanisms of such therapeutic interventions.

## **6. Conclusions**

The studies reviewed in this paper highlight that verbal instructions can have a tremendous impact on fear learning, both by directly installing fear and by moderating fear installed through stimulus pairings. We argue here that the results of these studies fit better with models that propose that conditioned fear is the result of a single process that generates expectancies about when certain antecedent stimuli might be followed by an aversive event. Nonetheless, more theoretical development is required to determine how different models of fear learning can be distinguished and more research is necessary to determine whether consistent qualitative and quantitative differences exist between the different pathways of fear learning. Finally, we highlight a number of future directions for the research on verbal instructions in fear conditioning, such as the further mapping of its procedural properties, investigating the role of factors such as experimental demand, believability and imagery, and relating the studies on this topic to the etiology and treatment of fear and anxiety disorders.

### **Acknowledgements**

Gaëtan Mertens and Iris Engelhard are supported by a VICI grant (453-15-005) awarded to Iris Engelhard by the Netherlands Organization for Scientific Research. Yannick Boddez is supported by a grant of the KU Leuven Research Council (PF/10/005; awarded to Dirk Hermans) and by a grant of the Belgian Science Policy Office (IUAPVII/33; KU Leuven PI: Tom Beckers). Jan De Houwer is supported by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (IUAPVII/33) and by Ghent University Methusalem Grant BOF16/MET\_V/002.

### References

- Ahmed, O., & Lovibond, P. F. (2015). The Impact of Instructions on Generalization of Conditioned Fear in Humans. *Behavior Therapy*, 46(5), 597–603. <https://doi.org/10.1016/j.beth.2014.12.007>
- Atlas, L. Y., & Phelps, E. A. (2018). Prepared stimuli enhance aversive learning without weakening the impact of verbal instructions. *Learning & Memory*, 25(2), 100–104. <https://doi.org/10.1101/lm.046359.117>
- Atlas, L. Y., Doll, B. B., Li, J., Daw, N. D., & Phelps, E. A. (2016). Instructed knowledge shapes feedback-driven aversive learning in striatum and orbitofrontal cortex, but not the amygdala. *eLife*, 5(MAY2016), 1–26. <https://doi.org/10.7554/eLife.15192>
- Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy*, 30(2), 133–142. [https://doi.org/10.1016/0005-7967\(92\)90136-5](https://doi.org/10.1016/0005-7967(92)90136-5)
- 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. <https://doi.org/10.3389/fpsyg.2014.00103>
- Bennett, M., Vervoort, E., Boddez, Y., Hermans, D., & Baeyens, F. (2015). Perceptual and conceptual similarities facilitate the generalization of instructed fear. *Journal of Behavior Therapy and Experimental Psychiatry*, 48, 149–155. <https://doi.org/10.1016/j.jbtep.2015.03.011>
- Blechert, J., Wilhelm, F. H., Williams, H., Braams, B. R., Jou, J., & Gross, J. J. (2015). Reappraisal facilitates extinction in healthy and socially anxious individuals. *Journal of Behavior Therapy and Experimental Psychiatry*, 46, 141–150. <https://doi.org/10.1016/j.jbtep.2014.10.001>
- 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 source of danger reduces the threat value of surrounding stimuli. *Journal of Experimental Psychopathology*, 4(4), 315–324. <https://doi.org/10.5127/jep.028512>
- Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2012). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry*, 44(2), 201–6. <https://doi.org/10.1016/j.jbtep.2012.08.003>
- Boddez, Y., Bennett, M. P., van Esch, S., & Beckers, T. (2017). Bending rules: the shape of the perceptual generalisation gradient is sensitive to inference rules. *Cognition and Emotion*, 0(0), 1–9. <https://doi.org/10.1080/02699931.2016.1230541>

- Braem, S., De Houwer, J., Demanet, J., Yuen, K. S. L., Kalisch, R., & Brass, M. (2017). Pattern analyses reveal separate experience-based fear memories in the human right amygdala. *The Journal of Neuroscience*, 908–17. <https://doi.org/10.1523/JNEUROSCI.0908-17.2017>
- Bridger, W. H., & Mandel, I. J. (1964). A comparison of GSR fear responses produced by threat and electric shock. *Journal of Psychiatric Research*, 2(1), 31–40. [https://doi.org/10.1016/0022-3956\(64\)90027-5](https://doi.org/10.1016/0022-3956(64)90027-5)
- Bridger, W. H., & Mandel, I. J. (1965). Abolition of the PRE by instructions in GSR conditioning. *Journal of Experimental Psychology*, 69(5), 476–482. <https://doi.org/10.1037/h0021764>
- Bublitzky, F., & Schupp, H. T. (2012). Pictures cueing threat: Brain dynamics in viewing explicitly instructed danger cues. *Social Cognitive and Affective Neuroscience*, 7(6), 611–622. <https://doi.org/10.1093/scan/nsr032>
- Büchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron*, 20(5), 947–957. [https://doi.org/10.1016/S0896-6273\(00\)80476-6](https://doi.org/10.1016/S0896-6273(00)80476-6)
- Carter, R. M., Hofstotter, C., Tsuchiya, N., & Koch, C. (2003). Working memory and fear conditioning. *Proceedings of the National Academy of Sciences*, 100(3), 1399–1404. <https://doi.org/10.1073/pnas.0334049100>
- Chan, C. K. Y., & Lovibond, P. F. (1996). Expectancy bias in trait anxiety. *Journal of Abnormal Psychology*, 105(4), 637–647. <https://doi.org/10.1037//0021-843X.105.4.637>
- Colgan, D. M. (1970). Effects of instructions on the skin resistance response. *Journal of Experimental Psychology*, 86(1), 108–112. <https://doi.org/10.1037/h0030011>
- Cook, S. W., & Harris, R. E. (1937). The verbal conditioning of the galvanic skin reflex. *Journal of Experimental Psychology*, 21(2), 202–210. <https://doi.org/10.1037/h0063197>
- Costa, V. D., Bradley, M. M., & Lang, P. J. (2015). From threat to safety: Instructed reversal of defensive reactions. *Psychophysiology*, 52(3), 325–332. <https://doi.org/10.1111/psyp.12359>
- Dadds, M. R., Bovbjerg, D. H., Redd, W. H., & Cutmore, T. R. H. (1997). Imagery in human classical conditioning. *Psychological Bulletin*, 122(1), 89–103. <https://doi.org/10.1037/0033-2909.122.1.89>
- Davey, G. C. L. (1992). Classical conditioning and the acquisition of human fears and phobias: A review and synthesis of the literature. *Advances in Behaviour Research and Therapy*, 14(1), 29–66. [https://doi.org/10.1016/0146-6402\(92\)90010-L](https://doi.org/10.1016/0146-6402(92)90010-L)
- Davey, G. C. L., & Matchett, G. (1994). Unconditioned stimulus rehearsal and the retention and enhancement of differential “fear” conditioning: Effects of trait and state anxiety. *Journal of Abnormal Psychology*, 103(4), 708–718. <https://doi.org/10.1037/0021-843X.103.4.708>

- Dawson, M. E., Catania, J. J., Schell, A. M., & Grings, W. W. (1979). Autonomic classical conditioning as a function of awareness of stimulus contingencies. *Biological Psychology*, 9(1), 23-40. [https://doi.org/10.1016/0301-0511\(79\)90020-6](https://doi.org/10.1016/0301-0511(79)90020-6)
- Dawson, M. E., & Furedy, J. J. (1976). The Role of Awareness in Human Differential Autonomic Classical Conditioning: The Necessary-Gate Hypothesis. *Psychophysiology*, 13(1), 50–53. <https://doi.org/10.1111/j.1469-8986.1976.tb03336.x>
- Dawson, M. E., & Biferno, M. A. (1973). Concurrent measurement of awareness and electrodermal classical conditioning. *Journal of Experimental Psychology*, 101(1), 55. <http://doi.org/10.1037/h0035524>
- Dawson, M. E., & Reardon, P. (1969). Effects of facilitory and inhibitory sets on GSR conditioning and extinction. *Journal of Experimental Psychology*, 82(3), 462–466. <https://doi.org/10.1037/h0028362>
- De Houwer, J. (2011). Evaluative conditioning: A review of functional knowledge and mental process theories. In T. R. Schachtman & S. Reilly (Eds.), *Associative learning and conditioning theory. Human and non-human applications* (pp. 399- 416). Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199735969.003.0130>
- De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon : On the relation between evaluative conditioning , evaluative conditioning via instructions, and persuasion. *Social Cognition*, 34(5), 480–494. <https://doi.org/10.1521/soco.2016.34.5.480>
- De Houwer, J., Hughes, S., & Brass, M. (2017). Toward a unified framework for research on instructions and other messages: An introduction to the special section on the power of instructions. *Neuroscience & Biobehavioral Reviews*. <https://doi.org/10.1016/j.neubiorev.2017.04.020>
- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, 135(3), 347–368. <https://doi.org/10.1037/a0014211>
- Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural Circuitry Underlying the Regulation of Conditioned Fear and Its Relation to Extinction. *Neuron*, 59(5), 829–838. <https://doi.org/10.1016/j.neuron.2008.06.029>
- Deltomme, B., Mertens, G., Tibboel, H., & Braem, S. (2017). Instructed fear stimuli bias visual attention. *Acta Psychologica*. <https://doi.org/10.1016/j.actpsy.2017.08.010>
- 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(1), 614–624. <https://doi.org/10.1016/j.jbtep.2011.08.006>
- Dimberg, U. (1987). Facial reactions, autonomic activity and experienced emotion: A three component model of emotional conditioning. *Biological psychology*, 24(2), 105-122. [https://doi.org/10.1016/0301-0511\(87\)90018-4](https://doi.org/10.1016/0301-0511(87)90018-4)

- Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, 62(3), 331–351. <https://doi.org/10.1901/jeab.1994.62-331>
- Drummond, P., White, K., & Ashton, R. (1978). Imagery vividness affects habituation rate. *Psychophysiology*, 15(3), 193–195. <https://doi.org/10.1111/j.1469-8986.1978.tb01362.x>
- Duits, P., Richter, J., Baas, J. M. P., Engelhard, I. M., Limberg-Thiesen, A., Heitland, I., ... Cath, D. C. (2017). Enhancing Effects of Contingency Instructions on Fear Acquisition and Extinction in Anxiety Disorders. *Journal of Abnormal Psychology*. <https://doi.org/10.1037/abn0000266>
- Dunsmoor, J. E., Martin, A., & LaBar, K. S. (2012). Role of conceptual knowledge in learning and retention of conditioned fear. *Biological Psychology*, 89(2), 300–305. <https://doi.org/10.1016/j.biopsycho.2011.11.002>
- Dymond, S., Schlund, M. W., Roche, B., De Houwer, J., & Freegard, G. P. (2012). Safe From Harm: Learned, Instructed, and Symbolic Generalization Pathways of Human Threat-Avoidance. *PLoS ONE*, 7(10). <https://doi.org/10.1371/journal.pone.0047539>
- Eifert, G. H. (1984). The effects of language conditioning on various aspects of anxiety. *Behaviour Research and Therapy*, 22(1), 13–21. [https://doi.org/10.1016/0005-7967\(84\)90028-7](https://doi.org/10.1016/0005-7967(84)90028-7)
- Eifert, G. H., & Schermelleh, K. (1985). Language conditioning, emotional instructions, and cognitions in conditioned responses to fear-relevant and fear-irrelevant stimuli. *Journal of Behavior Therapy and Experimental Psychiatry*, 16(2), 101–109. [https://doi.org/10.1016/0005-7916\(85\)90044-8](https://doi.org/10.1016/0005-7916(85)90044-8)
- Esteves, F., Dimberg, U., & Öhman, A. (1994). Automatically elicited fear: Conditioned skin conductance responses to masked facial expressions. *Cognition & Emotion*, 8(5), 393–413. <https://doi.org/10.1080/02699939408408949>
- Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias? *Clinical Psychology Review*, 26(7), 857–875. <https://doi.org/10.1016/j.cpr.2005.05.010>
- Field, A. P., Argyris, N. G., & Knowles, K. A. (2001). Who's afraid of the big bad wolf: a prospective paradigm to test Rachman's indirect pathways in children. *Behaviour Research and Therapy*, 39(11), 1259–1276. [https://doi.org/10.1016/S0005-7967\(00\)00080-2](https://doi.org/10.1016/S0005-7967(00)00080-2)
- Field, A. P., & Lawson, J. (2003). Fear information and the development of fears during childhood: Effects on implicit fear responses and behavioural avoidance. *Behaviour Research and Therapy*, 41(11), 1277–1293. [https://doi.org/10.1016/S0005-7967\(03\)00034-2](https://doi.org/10.1016/S0005-7967(03)00034-2)



- Field, A. P., & Schorah, H. (2007). The verbal information pathway to fear and heart rate changes in children. *Journal of Child Psychology and Psychiatry*, 48(11), 1088–1093. <https://doi.org/10.1111/j.1469-7610.2007.01772.x>
- Field, A. P., & Storksen-Coulson, H. (2007). The interaction of pathways to fear in childhood anxiety: A preliminary study. *Behaviour Research and Therapy*, 45(12), 3051–3059. <https://doi.org/10.1016/j.brat.2007.09.001>
- Freeman, G. L. (1930). The Galvanic Phenomenon and Conditioned Responses. *The Journal of General Psychology*, 3(4), 529–539. <https://doi.org/10.1080/00221309.1930.9918228>
- Fuhrer, M. J., & Baer, P. E. (1969). Cognitive Processes in Differential GSR Conditioning: Effects of a Masking Task. *The American Journal of Psychology*, 82(2), 168. <https://doi.org/10.2307/1421240>
- Gao, Y., Raine, A., Venables, P. H., Dawson, M. E., & Mednick, S. A. (2010). The development of skin conductance fear conditioning in children from ages 3 to 8 years. *Developmental science*, 13(1), 201–212. <https://doi.org/10.1111/j.1467-7687.2009.00874.x>
- Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *The Quarterly Journal of Experimental Psychology*, 65(9), 1657–1674. <https://doi.org/10.1080/17470218.2012.665061>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Grillon, C. (2009). D-Cycloserine Facilitation of Fear Extinction and Exposure-Based Therapy Might Rely on Lower-Level, Automatic Mechanisms. *Biological Psychiatry*, 66(7), 636–641. <https://doi.org/10.1016/j.biopsych.2009.04.017>
- Grillon, C., & Ameli, R. (1998). Effects of threat and safety signals on startle during anticipation of aversive shocks, sounds, or airblasts. *Journal of Psychophysiology*, 12(4), 329–337.
- Grillon, C., Ameli, R., Foot, M., & Davis, M. (2002). Fear-Potentiated Startle: Relationship to the Level of State/Trait Anxiety in Healthy Subjects. *Biological Psychiatry*, 33(8–9), 1–9. [https://doi.org/10.1016/0006-3223\(93\)90094-T](https://doi.org/10.1016/0006-3223(93)90094-T)
- Grillon, C., Ameli, R., Merikangas, K., Woods, S. W., & Davis, M. (1993). Measuring the time course of anticipatory anxiety using the fear-potentiated startle reflex. *Psychophysiology*, 30(4), 340–346. <https://doi.org/10.1111/j.1469-8986.1993.tb02055.x>
- Grillon, C., Ameli, R., Woods, S. W., Merikangas, K., & Davis, M. (1991). Fear-Potentiated Startle in Humans: Effects of Anticipatory Anxiety on the Acoustic Blink Reflex. *Psychophysiology*, 28(5), 588–595. <https://doi.org/10.1111/j.1469-8986.1991.tb01999.x>
- Grings, W. W. (1973). Cognitive factors in electrodermal conditioning. *Psychological Bulletin*, 79(3), 200–210. <https://doi.org/10.1037/h0033883>

- Grings, W. W., & Kimmel, H. D. (1959). Compound stimulus transfer for different sense modalities. *Psychological Reports*, 5(3), 253–260. <https://doi.org/10.2466/PR0.5.3.253-260>
- Grings, W. W., Schell, A. M., & Carey, C. A. (1973). Verbal control of an autonomic response in a cue reversal situation. *Journal of Experimental Psychology*, 99(2), 215–221. <https://doi.org/10.1037/h0034653>
- Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014). A review on human reinstatement studies: an overview and methodological challenges. *Learning & Memory*, 21(9), 424–440. <https://doi.org/10.1101/lm.036053.114>
- Hamm, A. O., & Weike, A. I. (2005). The neuropsychology of fear learning and fear regulation. *International Journal of Psychophysiology*, 57(1), 5–14. <https://doi.org/10.1016/j.ijpsycho.2005.01.006>
- Harvey, B., & Wickens, D. D. (1971). Effect of instructions on responsiveness to the CS and to the UCS in GSR conditioning. *Journal of Experimental Psychology*, 87(1), 137–140. <https://doi.org/10.1037/h0020864>
- Hermans, D., & Baeyens, F. (2013). Generalization as a basis for emotional change: Perceptual and non-perceptual processes. In D. Hermans, B. Rimé, & B. Mesquita (Eds.), *Changing emotions* (pp. 67–73). Hove, UK: Psychology Press.
- Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in Human Fear Conditioning. *Biological Psychiatry*, 60(4), 361–368. <https://doi.org/10.1016/j.biopsych.2005.10.006>
- Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van Den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy*, 43(4), 533–551. <https://doi.org/10.1016/j.brat.2004.03.013>
- Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy*, 40(3), 217–234. [https://doi.org/10.1016/S0005-7967\(01\)00006-7](https://doi.org/10.1016/S0005-7967(01)00006-7)
- Hill, F. A. (1967). Effects of instructions and subject's need for approval on the conditioned galvanic skin response. *Journal of Experimental Psychology*, 73(3), 461–467. <https://doi.org/10.1037/h0024263>
- Hinchy, J., Lovibond, P. F., & Ter-Horst, K. M. (1995). Blocking in human electrodermal conditioning. *The Quarterly Journal of Experimental Psychology Section B*, 48(1), 2–12. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7740123>
- Hosoba, T., Iwanaga, M., & Seiwa, H. (2001). The effect of UCS inflation and deflation procedures on “fear” conditioning. *Behaviour Research and Therapy*, 39(4), 465–475. [https://doi.org/10.1016/S0005-7967\(00\)00025-5](https://doi.org/10.1016/S0005-7967(00)00025-5)

- Hugdahl, K. (1978). Electrodermal conditioning to potentially phobic stimuli: effects of instructed extinction. *Behaviour Research and Therapy*, 16(5), 315–321.  
[https://doi.org/10.1016/0005-7967\(78\)90001-3](https://doi.org/10.1016/0005-7967(78)90001-3)
- Hugdahl, K., & Öhman, A. (1977). Effects of instruction on acquisition and extinction of electrodermal responses to fear-relevant stimuli. *Journal of Experimental Psychology: Human Learning & Memory*, 3(5), 608–618. <https://doi.org/10.1037/0278-7393.3.5.608>
- Javanbakht, A., Duval, E. R., Cisneros, M. E., Taylor, S. F., Kessler, D., & Liberzon, I. (2016). Instructed fear learning, extinction, and recall: additive effects of cognitive information on emotional learning of fear. *Cognition and Emotion*, 9931(May), 1–8.  
<https://doi.org/10.1080/02699931.2016.1169997>
- Jones, T., & Davey, G. C. L. (1990). The effects of cued UCS rehearsal on the retention of differential “fear” conditioning: An experimental analogue of the “worry” process. *Behaviour Research and Therapy*, 28(2), 159–164. [https://doi.org/10.1016/0005-7967\(90\)90028-H](https://doi.org/10.1016/0005-7967(90)90028-H)
- Joos, E., Vansteenwegen, D., & Hermans, D. (2012a). 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(2), 737–744.  
<https://doi.org/10.1016/j.jbtep.2011.10.011>
- Joos, E., Vansteenwegen, D., & Hermans, D. (2012b). Repetitive thought about an aversive learning experience maintains conditioned responding. *Journal of Experimental Psychopathology*, 3(1), 84–102. <http://doi.org/10.5127/jep.020811>
- Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and Aversive Behavior*. New York: Appleton-Century-Crofts.
- Kimmel, H. D., & Pennypacker, H. S. (1963). Differential GSR conditioning as a function of the CS-UCS interval. *Journal of Experimental Psychology*, 65(6), 559–563.  
<https://doi.org/10.1037/h0040417>
- King, N. J., Eleonora, G., & Ollendick, T. H. (1998). Etiology of childhood phobias: Current status of Rachman’s three pathways theory. *Behaviour Research and Therapy*, 36(3), 297–309. [https://doi.org/10.1016/S0005-7967\(98\)00015-1](https://doi.org/10.1016/S0005-7967(98)00015-1)
- Koster, E., Crombez, G., Van Damme, S., Verschuere, B., & De Houwer, J. (2005). Signals for threat modulate attentional capture and holding: Fear-conditioning and extinction during the exogenous cueing task. *Cognition & Emotion*, 19(5), 771–780.  
<https://doi.org/10.1080/02699930441000418>
- Lang, P. J. (1968). Fear reduction and fear behavior: Problems in treating a construct. In J. M. Schlien (Ed.), *Research in psychotherapy* (pp. 90–103). Washington, DC: American Psychological Association.

- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International Affective Picture System (IAPS): Technical Manual and Affective Ratings.
- LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences*, 111(8), 2871–2878. <https://doi.org/10.1073/pnas.1400335111>
- LeDoux, J. E., & Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: a two-system framework. *American Journal of Psychiatry*, 173(11), 1083–1093. <https://doi.org/10.1176/appi.ajp.2016.16030353>
- Lipp, O. V., 2006. Human fear learning: contemporary procedures and measurement. In: Craske, M. G., Hermans, D., Vansteenwegen, D. (Eds.), *Fear and Learning: From Basic Processes to Clinical Implications* (pp. 37–52). American Psychological Association, Washington, DC. <https://doi.org/10.1037/11474-002>
- Lipp, O. V., & Vaitl, D. (1990). Reaction time task as unconditional stimulus - Comparing aversive and nonaversive unconditional stimuli. *The Pavlovian Journal of Biological Science*, 25(2), 77–83. <https://doi.org/10.1007/BF02964606>
- 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(5), 678–687. <https://doi.org/10.1016/j.brat.2008.02.005>
- Lissek, S., Orme, K., McDowell, D. J., Johnson, L. L., Luckenbaugh, D. A., Baas, J. M., ... Grillon, C. (2007). Emotion regulation and potentiated startle across affective picture and threat-of-shock paradigms. *Biological Psychology*, 76(1–2), 124–133. <https://doi.org/10.1016/j.biopsycho.2007.07.002>
- Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., ... Merz, C. J. (2017). Don't fear "fear conditioning": Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, 77, 247–285. <https://doi.org/10.1016/j.neubiorev.2017.02.026>
- Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 97–106. <https://doi.org/10.1037/0278-7393.29.1.97>
- Lovibond, P. F. (2004). Cognitive processes in extinction. *Learning & Memory*, 11(5), 495–500. <https://doi.org/10.1101/lm.79604>
- Lovibond, P. F. (2011). Learning and anxiety: A cognitive perspective. In T. R. Schachtman & S. Reilly (Eds.), *Associative Learning and Conditioning: Human and non-human Applications* (pp. 104–120). New York: Oxford University Press.
- 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(8), 716–720.  
<https://doi.org/10.1016/j.brat.2009.04.013>
- Luck, C. C., & Lipp, O. V. (2015a). A potential pathway to the relapse of fear? Conditioned negative stimulus evaluation (but not physiological responding) resists instructed extinction. *Behaviour Research and Therapy*, 66, 18–31. <https://doi.org/10.1016/j.brat.2015.01.001>
- Luck, C. C., & Lipp, O. V. (2015b). To remove or not to remove? Removal of the unconditional stimulus electrode does not mediate instructed extinction effects. *Psychophysiology*, 52(9), 1248–1256. <https://doi.org/10.1111/psyp.12452>
- Luck, C. C., & Lipp, O. V. (2016a). Instructed extinction in human fear conditioning: History, recent developments, and future directions. *Australian Journal of Psychology*, 68(3), 209–227. <https://doi.org/10.1111/ajpy.12135>
- Luck, C. C., & Lipp, O. V. (2016b). The influence of contingency reversal instructions on electrodermal responding and conditional stimulus valence evaluations during differential fear conditioning. *Learning and Motivation*, 54, 1–11.  
<https://doi.org/10.1016/j.lmot.2016.05.001>
- Luck, C. C., & Lipp, O. V. (2018). Verbal instructions targeting valence alter negative conditional stimulus evaluations (but do not affect reinstatement rates). *Cognition and Emotion*, 32(1), 61–80. <https://doi.org/10.1080/02699931.2017.1280449>
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, 95(1), 15. <http://doi.org/10.1037/0021-843X.95.1.15>
- Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, 15(2), 197–237. [https://doi.org/10.1016/0010-0285\(83\)90009-9](https://doi.org/10.1016/0010-0285(83)90009-9)
- McLaren, I. P. L., Forrest, C. L. D., McLaren, R. P., Jones, F. W., Aitken, M. R. F., & Mackintosh, N. J. (2014). Associations and propositions: The case for a dual-process account of learning in humans. *Neurobiology of Learning and Memory*, 108, 185–195.  
<https://doi.org/10.1016/j.nlm.2013.09.014>
- McNally, R. J. (1981). Phobias and preparedness: instructional reversal of electrodermal conditioning to fear-relevant stimuli. *Psychological Reports*, 48(1), 175–180.  
<https://doi.org/10.2466/pr0.1981.48.1.175>
- McNally, R. J., & Foa, E. B. (1986). Preparedness and resistance to extinction to fear-relevant stimuli: A failure to replicate. *Behaviour Research and Therapy*, 24(5), 529–535.  
[https://doi.org/10.1016/0005-7967\(86\)90033-1](https://doi.org/10.1016/0005-7967(86)90033-1)
- McNally, R. J. (1987). Preparedness and phobias: A review. *Psychological Bulletin*, 101(2), 283.  
<http://doi.org/10.1037/0033-2909.101.2.283>

- Mechias, M. L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. *NeuroImage*, 49(2), 1760–1768. <https://doi.org/10.1016/j.neuroimage.2009.09.040>
- Mertens, G., & De Houwer, J. (2016a). Potentiation of the startle reflex is in line with contingency reversal instructions rather than the conditioning history. *Biological Psychology*, 113, 91–99. <https://doi.org/10.1016/j.biopsycho.2015.11.014>
- Mertens, G., & De Houwer, J. (2016b). 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. <https://doi.org/10.1016/j.jbtep.2015.11.001>
- Mertens, G., & De Houwer, J. (2017). Can threat information bias fear learning? Some tentative results and methodological considerations. *Journal of Experimental Psychopathology*. <https://doi.org/10.5127/jep.060616>
- Mertens, G., Kuhn, M., Raes, A. K., Kalisch, R., De Houwer, J., & Lonsdorf, T. B. (2016). Fear expression and return of fear following threat instruction with or without direct contingency experience. *Cognition and Emotion*, 30(5), 968–984. <https://doi.org/10.1080/02699931.2015.1038219>
- Mertens, G., Raes, A. K., & De Houwer, J. (2016). Can prepared fear conditioning result from verbal instructions? *Learning and Motivation*, 53, 7–23. <https://doi.org/10.1016/j.lmot.2015.11.001>
- Mineka, S., & Öhman, A. (2002). Phobias and preparedness: The selective, automatic, and encapsulated nature of fear. *Biological Psychiatry*, 52(10), 927–937. [https://doi.org/10.1016/S0006-3223\(02\)01669-4](https://doi.org/10.1016/S0006-3223(02)01669-4)
- Miskovic, V., & Keil, A. (2012). Acquired fears reflected in cortical sensory processing: A review of electrophysiological studies of human classical conditioning. *Psychophysiology*, 49(9), 1230–1241. <https://doi.org/10.1111/j.1469-8986.2012.01398.x>
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *The Behavioral and Brain Sciences*, 32(2), 183-98-246. <https://doi.org/10.1017/S0140525X09000855>
- Morriss, J., Christakou, A., & van Reekum, C. M. (2015). Nothing is safe: Intolerance of uncertainty is associated with compromised fear extinction learning. *Biological Psychology*, (October). <https://doi.org/10.1016/j.biopsycho.2016.05.001>
- Muris, P., & Field, A. P. (2010). The role of verbal threat information in the development of childhood fear. “Beware the Jabberwock!” *Clinical Child and Family Psychology Review*, 13(2), 129–150. <https://doi.org/10.1007/s10567-010-0064-1>
- Muris, P., Rassin, E., Mayer, B., Smeets, G., Huijding, J., Remmerswaal, D., & Field, A. (2009). Effects of verbal information on fear-related reasoning biases in children. *Behaviour Research and Therapy*, 47(3), 206–214. <https://doi.org/10.1016/j.brat.2008.12.002>

- Neumann, D. L. (2007). The resistance of renewal to instructions that devalue the role of contextual cues in a conditioned suppression task with humans. *Learning and Motivation*, 38(2), 105–127. <https://doi.org/10.1016/j.lmot.2006.11.002>
- Norrholm, S. D., Jovanovic, T., Vervliet, B., Myers, K. M., Davis, M., Rothbaum, B. O., & Duncan, E. J. (2006). Conditioned fear extinction and reinstatement in a human fear-potentiated startle paradigm. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 13(6), 681–5. <https://doi.org/10.1101/lm.393906>
- Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, 108(3), 483–522. <https://doi.org/10.1037//0033-295X.108.3.483>
- Olsson, A., & Phelps, E. a. (2007). Social learning of fear. *Nature Neuroscience*, 10(9), 1095–102. <https://doi.org/10.1038/nn1968>
- Olsson, A., & Phelps, E. A. (2004). Learned fear of “unseen” faces after Pavlovian, observational, and instructed fear. *Psychological Science*, 15(12), 822–828. <https://doi.org/10.1111/j.0956-7976.2004.00762.x>
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776–783. <http://dx.doi.org/10.1037/h0043424>
- Phelps, E. a, O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., & Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. *Nature Neuroscience*, 4(4), 437–441. <https://doi.org/10.1038/86110> [pii]
- Rachman, S. (1977). The conditioning theory of fear acquisition: A critical examination. *Behaviour Research and Therapy*, 15(5), 375–387. [https://doi.org/10.1016/0005-7967\(77\)90041-9](https://doi.org/10.1016/0005-7967(77)90041-9)
- Rachman, S. (1991). Neo-conditioning and the classical theory of fear acquisition. *Clinical Psychology Review*, 11(2), 155–173. [https://doi.org/10.1016/0272-7358\(91\)90093-A](https://doi.org/10.1016/0272-7358(91)90093-A)
- Raes, A. K., De Houwer, J., De Schryver, M., Brass, M., & Kalisch, R. (2014). Do CS-US pairings actually matter? A within-subject comparison of instructed fear conditioning with and without actual CS-US pairings. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0084888>
- 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(3), 212–218. <https://doi.org/10.1016/j.brat.2010.12.004>
- Reiss, S. (1980). Pavlovian conditioning and human fear: An expectancy model. *Behavior Therapy*, 11(3), 380–396. [https://doi.org/10.1016/S0005-7894\(80\)80054-2](https://doi.org/10.1016/S0005-7894(80)80054-2)

- Remmerswaal, D., Huijding, J., Bouwmeester, S., Brouwer, M., & Muris, P. (2014). Cognitive bias in action: Evidence for a reciprocal relation between confirmation bias and fear in children. *Journal of Behavior Therapy and Experimental Psychiatry*, 45(1), 26–32. <https://doi.org/10.1016/j.jbtep.2013.07.005>
- Scheveneels, S., Boddez, Y., Bennett, M. P., & 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. <https://doi.org/10.1016/j.jbtep.2017.03.002>
- Schindler, B., Vriends, N., Margraf, J., & Stieglitz, R. D. (2016). Ways of acquiring flying phobia. *Depression and Anxiety*, 33(2), 136–142. <https://doi.org/10.1002/da.22447>
- Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy*, 2(3), 307–320. [https://doi.org/10.1016/S0005-7894\(71\)80064-3](https://doi.org/10.1016/S0005-7894(71)80064-3)
- Sevenster, D., Beckers, T., & Kindt, M. (2012). Instructed extinction differentially affects the emotional and cognitive expression of associative fear memory. *Psychophysiology*, 49(10), 1426–1435. <https://doi.org/10.1111/j.1469-8986.2012.01450.x>
- Shanks, D. R. (2010). Learning: from association to cognition. *Annual Review of Psychology*, 61, 273–301. <https://doi.org/10.1146/annurev.psych.093008.100519>
- Shurick, A. A., Hamilton, J. R., Harris, L. T., Roy, A. K., Gross, J. J., & Phelps, E. A. (2012). Durable effects of cognitive restructuring on conditioned fear. *Emotion*, 12(6), 1393–1397. <https://doi.org/10.1037/a0029143>
- Soeter, M., & Kindt, M. (2012). Erasing fear for an imagined threat event. *Psychoneuroendocrinology*, 37(11), 1769–1779. <https://doi.org/10.1016/j.psyneuen.2012.03.011>
- Spence, K. W., & Runquist, W. N. (1958). Temporal effects of conditioned fear on the eyelid reflex. *Journal of Experimental Psychology*, 55(6), 613–616. <https://doi.org/10.1037/h0049214>
- Struyf, D., Zaman, J., Vervliet, B., & Van Diest, I. (2015). Perceptual discrimination in fear generalization: Mechanistic and clinical implications. *Neuroscience and Biobehavioral Reviews*, 59, 201–207. <https://doi.org/10.1016/j.neubiorev.2015.11.004>
- Swenson, R. P., & Hill, F. A. (1970). Effects of instruction and interstimulus interval in human GSR conditioning. *Psychonomic Science*, 21(6), 369–370. <https://doi.org/10.3758/BF03335824>
- Switzer, C. A. (1933). Disinhibition of the Conditioned Galvanic Skin Response. *The Journal of General Psychology*, 9(1), 77–100. <https://doi.org/10.1080/00221309.1933.9920914>
- Tabbert, K., Merz, C. J., Klucken, T., Schweckendiek, J., Vaitl, D., Wolf, O. T., & Stark, R. (2011). Influence of contingency awareness on neural, electrodermal and evaluative



- responses during fear conditioning. *Social Cognitive and Affective Neuroscience*, 6(4), 495–506. <https://doi.org/10.1093/scan/nsq070>
- Tomarken, A. J., Mineka, S., & Cook, M. (1989). Fear-relevant selective associations and covariation bias. *Journal of Abnormal Psychology*, 98(4), 381. <http://doi.org/10.1037/0021-843X.98.4.381>
- Ugland, C. C. O., Dyson, B. J., & Field, A. P. (2013). An ERP study of the interaction between verbal information and conditioning pathways to fear. *Biological Psychology*, 92(1), 69–81. <https://doi.org/10.1016/j.biopsycho.2012.02.003>
- 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(1), 51–58. <https://doi.org/10.1016/j.biopsycho.2012.01.006>
- Vervliet, B., Kindt, M., Vansteenwegen, D., & Hermans, D. (2010). Fear generalization in humans: Impact of verbal instructions. *Behaviour Research and Therapy*, 48(1), 38–43. <https://doi.org/10.1016/j.brat.2009.09.005>
- Warren, V. T., Anderson, K. M., Kwon, C., Bosshardt, L., Jovanovic, T., Bradley, B., & Norrholm, S. D. (2014). Human fear extinction and return of fear using reconsolidation update mechanisms: The contribution of on-line expectancy ratings. *Neurobiology of Learning and Memory*, 113(OCTOBER), 165–173. <https://doi.org/10.1016/j.nlm.2013.10.014>
- Weymar, M., Bradley, M. M., Hamm, A. O., & Lang, P. J. (2013). When fear forms memories: Threat of shock and brain potentials during encoding and recognition. *Cortex*, 49(3), 819–826. <https://doi.org/10.1016/j.cortex.2012.02.012>
- White, K., & Davey, G. C. L. (1989). Sensory preconditioning and UCS inflation in human “fear” conditioning. *Behaviour Research and Therapy*, 27(2), 161–166. [https://doi.org/10.1016/0005-7967\(89\)90074-0](https://doi.org/10.1016/0005-7967(89)90074-0)
- Wilson, G. D. (1968). Reversal of differential GSR conditioning by instructions. *Journal of Experimental Psychology*, 76(3, Pt.1), 491–493. <https://doi.org/10.1037/h0025540>
- Yaremko, R. M., & Werner, M. (1974). Cognitive conditioning: Imagined stimulus contiguity and the third interval conditional GSR. *Pavlovian Journal of Biological Science*, 9(4), 215–221. <https://doi.org/10.1007/BF03001505>
- Zeng, Q., Jia, Y., Wang, Y., Zhang, J., Liu, C., & Zheng, X. (2015). Retrospective reversal of extinction of conditioned fear by instruction. *Consciousness and Cognition*, 35, 171–177. <https://doi.org/10.1016/j.concog.2015.05.011>
- 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(October), 270. <https://doi.org/10.3389/fnbeh.2015.00270>