Instructed Extinction in Human Fear Conditioning: History, Recent Developments and Future Directions

Camilla C. Luck[1, 2] & Ottmar V. Lipp[1, 2]

[1]School of Psychology and Speech Pathology, Curtin University, Australia

[2]ARC-SRI: Science of Learning Research Centre

Author Notes

Address for correspondence:

Camilla C. Luck, School of Psychology and Speech Pathology, Curtin University, GPO Box U1987 Perth WA 6845, Australia. Email: c.luck@curtin.edu.au

Acknowledgements

This work was supported by grants number DP120100750 and SR120300015 from the Australian Research Council.

Word count: 11730

Abstract

Instructed extinction is an experimental manipulation which involves informing participants after the acquisition of fear learning that the unconditional stimulus will no longer be presented. It has been used as a laboratory analogue to assess the capacity of cognitive interventions to reduce experimentally induced fear. In this review we examine and integrate research on instructed extinction and discuss its implications for clinical practice. Overall, the results suggest that instructed extinction reduces conditional fear responding and facilitates extinction learning, except when conditional stimulus valence is assessed as an index of fear or when fear is conditioned to images of animal fear relevant stimuli (snakes and spiders) or with a very intense unconditional stimulus. These exceptions highlight potential boundary conditions for the reliance on cognitive interventions when treating fear in clinical settings.

Key words: Instructed extinction; fear conditioning; cognitive interventions; return of fear; anxiety.

Fear can be a learned response – a neutral stimulus will elicit fear independently if it has been associated with an aversive stimulus. There are a number of pathways in which this fear association can be formed – including repeated pairings between the neutral and the aversive stimulus (experiential learning); observing another individual displaying fear to the neutral stimulus (observational learning); or being informed that the neutral stimulus is predictive of the aversive event (informational learning) (Rachman, 1968; Rachman, 1977). If contained, fear is adaptive as it facilitates defensive responding allowing the escape from, or avoidance of, dangerous situations, but if fear becomes exaggerated or is not appropriately regulated, it can develop into an anxiety disorder (Quinn & Fanselow, 2006). Anxiety disorders are emotionally and economically costly and will affect 25% of the population during their lifetime (Kessler, Koretz, Merikangas, & Wang, 2004).

Developing treatments which are efficacious in both the short, and the long term, has become a central focus of research on anxiety disorders. The short term success of gold standard treatments is well documented (Bisson & Andrew, 2007; Ougrin, 2011; Sánchez-Meca, Rosa-Alcàzar, Marín-Martinez, & Gómez-Conesa, 2010), but one to two thirds of these successfully treated patients will relapse within eight years (Craske, 1999). This clinical observation is consistent with results of laboratory research showing that fear extinction does not erase the original fear memory, but instead lays down a new context specific extinction memory (Bouton, 2002). After extinction learning, the original fear memory often re-emerges resulting in the return of fear (Rachman, 1966; for a review see Vervliet, Hermans & Craske, 2013). Understanding why fear re-emerges and how this phenomenon can be reduced in the laboratory is crucial to developing long lasting treatments.

Common anxiety treatments and their effects on fear and fear relapse can be modelled in the controlled laboratory environment (Craske, Hermans, & Vansteenwegen, 2006). Instructed extinction is a laboratory manipulation which involves using instructions to break the association between the neutral stimulus and the aversive stimulus (Luck & Lipp, 2015a). It is often considered a laboratory analogue for a cognitive intervention and has been used in a number of different contexts and under a number of different names over the last 60 years. In this review we will give a brief overview of the paradigms and measures involved in instructed extinction research before examining the research conducted with this manipulation within the human fear conditioning paradigm. After the review of the literature, we will integrate the findings, discuss their significance for clinical practice, and offer possible directions for future research.

A Brief Introduction to Human Fear Conditioning

Classical fear conditioning can be used to model the development, treatment, and relapse of human fear (Craske, Hermans, & Vansteenwegen, 2006). During classical fear acquisition, a neutral conditional stimulus (CS), e.g. a picture or tone, is repeatedly paired with an aversive unconditional stimulus (US), e.g. an electrotactile shock or loud noise. After repeated pairings, the CS becomes a signal for the US and elicits fear responding independently. During classical fear extinction, the CS is presented alone, and fear to the CS reduces. In the laboratory, the return of fear can be examined with three experimental manipulations. Spontaneous recovery, the return of fear after the mere passage of time, can be assessed by presenting the CS after a break in the experiment or after the participants have returned to the lab at a different time. Renewal, the return of fear after a context change, can be assessed by examining responding to the CS in a context that differs from the one used during extinction training; and reinstatement, the return of

fear after presentation of the aversive stimulus, can be measured by presenting the CS after unsignaled presentations of the US (Bouton, 2002; Vervliet, Hermans, & Craske, 2013).

Acquisition, extinction, and the return of fear can be assessed within two variations of the fear conditioning paradigm – single cue and differential fear conditioning. In a single cue design, participants are presented with one CS paired with the US, and their responding is compared with a control group who receive random, or explicitly unpaired, presentations of the CS and the US. The single cue design has been criticised as it does not control for orienting and other non-associative processes that may affect responding to the CS. Moreover, selecting the appropriate control is difficult and if an explicitly unpaired stimulus sequence is used it can result in inhibitory conditioning to the CS. A differential fear conditioning design embeds the control for non-associative factors into a within participants design by using two CSs, one paired with the US (CS+) and another presented alone (CS-) (Lipp, 2006).

A number of important factors which can influence conditioning vary across studies including the CS duration, the interval between the CS and the US (interstimulus interval; ISI), and the reinforcement rate (for a detailed discussion see Lipp, 2006). In delay conditioning, CS offset coincides with, or is preceded by, the onset of the US, whereas, in trace conditioning, there is a time interval between CS offset and US onset. Delay conditioning is usually acquired faster and is more robust than trace conditioning (see for instance Lipp, Siddle & Dall, 2003). The choice of CS duration largely depends on the measure used to index conditioning. If autonomic responses are to be measured long CS durations (typically 6 or 8 seconds) are usually used to separate the unconditional response elicited by the US from conditional responding to the CS. Shorter CS durations are acceptable if the response system used to index conditioning is quick (i.e. eye blink conditioning or self-report measures). The ISI is the duration between the onset of

the CS and the onset of the US and is dependent on both the CS duration and the interval between the CS offset and US onset. The reinforcement rate is the percentage of times that the CS is paired with the US during acquisition out of the total number of CS presentations.

Human fear learning can be assessed across three different response levels – physiologically, behaviourally, and verbally (Lang, 1985). The focus of human fear conditioning research has been on physiological and verbal indices and we will describe the common measures used in studies of instructed extinction in this section. Each measure used to index fear learning has advantages and limitations and therefore the effect of instructed extinction on human fear should be assessed across a number of different measures.

Electrodermal Responding

Electrodermal responding reflects variations in the conductivity of human skin to electrical currents due to changes in sympathetic nervous system activation of the eccrine sweat glands (Dawson, Schell, & Filion, 2007). It is the most frequently used measure in human fear conditioning and the most common index of instructed extinction. Electrodermal responding is sensitive to the psychological processes important during associative learning, such as orienting to, and the anticipation of, salient events. It is not selectively sensitive to fear learning, however, showing the same response pattern regardless of whether an aversive or a non-aversive US is used (Lipp and Vaitl, 1990). Electrodermal responding can be scored by distinguishing multiple response components during the CS-US interval or by scoring a single response during the entire interval. If a long CS duration is used, a first interval response will emerge within 1-4 seconds of CS onset and a second interval response will emerge within 4-7 seconds (6s ISI) or 4-9 seconds (8s ISI) of CS onset (Prokasy & Kumpfer, 1973). First interval responding is more sensitive to orienting elicited by CS onset and second interval responding is more sensitive to the

anticipation of the US (Öhman, 1983), however there is considerable covariation. The entire interval scoring technique scores the largest response occurring during the CS-US interval as a single index. Luck and Lipp (2016) compared multiple response scoring and entire interval scoring of data from an instructed extinction study and provided evidence that, because of a dissociation between orienting and anticipation, the instructed extinction effects which were detected using multiple response scoring were lost with entire interval scoring.

Heart Rate

Heart rate changes provide a cardiovascular index of conditioning and heart rate responses to a CS, in anticipation of a US, often consist of an initial deceleration, a transient acceleration, and a subsequent deceleration. The initial deceleration reflects orienting to the CS, whereas the second and third component reflect the anticipation of the US. Conditioned heart rate responses seem to be sensitive to the affective valence of the US, with the accelerative heart rate response component believed to reflect anticipation of an aversive stimulus as it is most prominent in studies using intense USs or fear relevant CSs (Lipp, 2006).

Blink Startle Responding

Blink startle responding is a skeletal nervous system measure of the brainstem startle reflex. It is not under cognitive control and is linearly modulated by valence, such that startle responding is inhibited if elicited during pleasant stimuli and potentiated if elicited during unpleasant stimuli (Lang, Bradley, & Cuthbert, 1990), but only if these stimuli are high in arousal (Cuthbert, Bradley, & Lang, 1996). Startle responding is considered a robust measure of fear learning and there are some reports that startle is potentiated only during anticipation of

aversive USs (Hamm & Vaitl, 1996). Others have argued that conditioning with aversive and non-aversive USs can elicit the same pattern of startle response modulation (Lipp et al., 2003).

Conditional Stimulus Valence

The addition of verbal measures of CS valence to conditioning designs has become popular due to the difficulties assessing valence reliably with physiological indices. CS valence can be assessed before and after conditioning training, or throughout conditioning (online) with a continuous response indicator (Lipp, 2006). Pre/post measures cannot index real-time changes in valence and may be confounded by renewal effects as they are frequently recorded in a different experimental context. In instructed extinction studies continuous assessments of CS valence are preferred as they can be obtained during the CS immediately after the instructed extinction manipulation, allowing for the assessment of instructed extinction effects before additional learning occurs (Luck & Lipp, 2015a).

Unconditional Stimulus Expectancy

US expectancy is measured to assess participants' anticipation of the US or awareness of the CS-US contingency. US expectancy is often assessed as a manipulation check after the completion of the experiment by asking participants to identify which stimulus had been associated with the US. Alternatively, US expectancy can be assessed as a dependent variable online throughout conditioning training (Lipp, 2006).

Instructed Extinction Manipulation

Instructed extinction is an experimental manipulation which assesses whether receiving instructions about the absence of the US is sufficient to reduce conditional responding. During instructed extinction, the experimenter interacts with participants after the last acquisition trial.

In the instruction group, participants are informed that the US will no longer be presented and the devices used to deliver the US (shock electrode or headphones) are often removed. Responding in the instruction group is then compared with a control group, who experience a similar interaction with the experimenter (i.e. to check the electrodes) but are not given information about the CS-US contingency. To allow for the identification, and possible exclusion, of participants who did not believe the instructions, the experimental group are typically asked whether they believed the instructions after the experiment.

Assessing instructed extinction effects relative to a control group who are exposed to the same level of interaction with the experimenter, but not instructed, controls for the effects of the manipulation on overall arousal and, potentially, conditional responding. The shock electrode is often removed to strengthen the manipulation and reduce the number of participants who do not believe the instructions. Some argue that this removal could reduce arousal levels and add a non-cognitive component to the manipulation. A direct comparison between instructed extinction with, and without shock electrode removal, however has failed to substantiate this concern (Luck & Lipp, 2015b). Generally two types of instruction effects can be assessed. Instructed extinction can abolish differential conditional responding on the very first trial of extinction or it can facilitate extinction learning. A reduction of conditional responding on the first trial of extinction in the instruction group, relative to the control group can be attributed to the provision of information alone. Facilitation of extinction learning can be considered an interactive effect between explicit extinction training and the instructional manipulation.

Instructed Extinction with Non-Fear Relevant Conditional Stimuli

Cook and Harris (1937) were the first to hypothesise that a conditional electrodermal response could be removed by breaking the CS-US association with verbal instructions. Using a

single cue short delay conditioning paradigm (3s ISI – US presented at CS offset; for further details of individual experiments see Table 1), participants were conditioned with a tone and an electrotactile shock throughout acquisition. After instructed extinction, electrodermal responding was considerably reduced in the instruction group in comparison with the non-instructed control group. Soon after, this initial observation was confirmed by Mowrer (1938) who reported that the conditional electrodermal response could be 'be switched on and off' by removing and reattaching the shock electrode or by using a buzzer system to indicate phases in which the US could be expected.

Notterman, Schoenfeld and Bersh (1952) extended this line of research by confirming that the conditional heart rate response was also subject to instructed extinction. During acquisition, participants were conditioned using a single cue trace conditioning design (7s ISI – 6s trace interval). Instructed extinction did not influence conditional heart rate responses within the first 5 extinction trials but extinction learning was facilitated in the instruction group during the last 5 extinction trials.

Sensitisation is a non-associative learning process in which the mere presentation of aversive stimuli can enhance electrodermal responding to neutral stimuli. Silverman (1960) argued that because the earlier instructed extinction studies did not include a pseudo-conditioning control group it was not clear whether instructed extinction was influencing a conditional response or a sensitised response. To confirm this, he compared the effect of instructed extinction on conditional electrodermal responding after three different acquisition procedures – conditioning with a 2.5s ISI (0.5s trace interval), conditioning with a 8s ISI (6s trace interval), or a pseudo-conditioning (unpaired) control group. Instructed extinction reduced electrodermal responding in the 2.5s ISI and the control group, but not in the 8s ISI group. The

reduction of electrodermal responding in the 2.5s ISI group confirmed that instructed extinction could reduce a conditional response, but failure to find instructed extinction effects using a 8s ISI is surprising especially in light of the significant reduction detected in the unpaired control group. Silverman suggested that the long trace interval could be anxiety arousing and protect against instructed extinction effects, but such an interpretation is not consistent with the results of Notterman et al. (1952) who also used a 6s trace interval.

Lindley and Moyer (1961) examined the effects of instructed extinction on the conditioned finger withdrawal response (conditional movement of the finger after electrotactile shock to the finger) after minimal and extended acquisition training. Participants were conditioned using a single cue short trace (1s ISI – 0.5s trace interval) conditioning paradigm. Consistent with research on electrodermal responding and heart rate, instructed extinction reduced the conditioned finger withdrawal response. There was also some evidence that this reduction was larger in the participants who received minimal acquisition training.

Wickens, Allen and Hill (1963) investigated whether US intensity could moderate the effect instructed extinction on the conditional electrodermal response. Using a single cue short delay conditioning paradigm (0.5s ISI – US presented at CS offset), participants were conditioned with a weak or a strong electrotactile shock. Instructed extinction did not influence conditional responding on the first extinction trial, but did facilitate the speed of extinction learning relative to the control group. No interactions between US intensity and instructed extinction were detected. This finding was confirmed by Grings and Lockhart (1963) who examined whether US intensity and amount of acquisition training would moderate the effect of instructed extinction on the conditional electrodermal response. Using a single cue long delay conditioning paradigm (5s ISI – US presented at CS offset) all participants viewed 3 CSs paired

with a different US intensity (high, medium, low). Half of the participants received 9 CS-US pairings (3 of each CS) and the other half received 36 CS-US pairings (12 of each CS). Instructed extinction reduced electrodermal responding on the first extinction trial of each CS, but was not influenced by US intensity or the number of CS-US pairing during acquisition.

Bridger and Mandel (1964) failed to find facilitation of extinction learning after instructed extinction in a long delay differential conditioning design (6s ISI – US delivered 1s before CS offset) using a painful electrotactile shock US. They hypothesised that conditional electrodermal responding established during CS-US pairings or during a threat of shock phase would be differentially sensitive to instructed extinction. During acquisition, both the conditioning and the threat group acquired differential responding which did not differ on the last acquisition trial. After instructed extinction, differential responding was eliminated in the threat group, but remained intact in the conditioning group. Bridger and Mandel suggest that instructed extinction will eliminate a conditional response which was established via instructions but not a conditional response which was established via direct CS-US pairings. This suggestion is not consistent with the majority of instructed extinction studies in the literature, but could occur because of the intense US that was used.

More consistent with prior research, Bridger and Mandel (1965) report that instructed extinction facilitated the extinction of a conditional electrodermal response established with direct CS-US pairings. Using a short delay differential conditioning design (0.5s ISI – US on CS+ offset), reinforcement rate during acquisition training was varied between groups. One group received acquisition training with a partial reinforcement schedule (25%) and another with a continuous reinforcement schedule (100%). The reinforcement schedule did not moderate the instruction effects. All groups (controls and instructions) showed continued differential

responding on the first extinction trial, but the magnitude of this differential response was reduced in the instruction groups and subsequent extinction learning was facilitated.

Mandel and Bridger (1967) examined the effect of instructed extinction after conditioning with three different acquisition procedures – a forward conditioning short (0.5s) delay group, a forward conditioning long (5s) delay group, and a backward conditioning group. During acquisition, all groups acquired differential responding between CS+ and CS-. During the first five extinction trials, differential responding was absent in the backward conditioning groups (control and instruction), but still present in all other groups. Differential responding was not present in any group during the last five extinction trials.

In the studies reported by Bridger and Mandel differential electrodermal responding was consistently present in the instruction groups during the first extinction trial and instructed extinction did not facilitate the speed of extinction learning in Bridger and Mandel (1965) or Mandel and Bridger (1967). These findings suggest that conditional electrodermal responding is not always eliminated immediately by instructed extinction. Mandel and Bridger (1973) suggest that strong instruction effects are not present in their studies because they used a very painful shock as the US. Wickens et al. (1963) and Grings and Lockhart (1963) have reported that US intensity does not moderate instructed extinction effects, however the maximum US intensity in these studies was set by the participant to be unpleasant but not painful. In contrast, participants in Bridger and Mandel's studies received a pre-set shock intensity that was perceived by all participants as very painful. Mandel and Bridger report that 10% of the participants refused to continue participation and that many indicated fear or anger about remaining in the experiment and assert that the mildly uncomfortable shock used in most prior studies would not permit the acquisition of conditional responses which are not merely reflections of cognitive expectancy.

Fuhrer and Baer (1980) aimed to examine whether resistance to instructed extinction could be obtained with a less noxious electrotactile shock and whether instructed extinction effects would differ between a 0.5s ISI and a 5s ISI (delay conditioning – US on CS+ offset). Throughout the experiment a continuous measure of US expectancy was assessed alongside electrodermal responding. All participants were informed after acquisition that the US would no longer be presented and participants were then divided into 'believers' and 'non-believers' based on their US expectancy. During the first extinction block (3 extinction trials), participants who reported not expecting the US continued to show differential responding between the CS+ and CS- in both ISI groups. A similar, but non-significant, differential pattern was detected in the participants who reported still expecting the US and differential responding was eliminated in all groups after the first extinction block. Fuhrer and Baer (1980) interpret their findings as a demonstration of conditional responding which is inconsistent with cognitive expectancies after conditioning with mildly unpleasant US, but this interpretation should be treated with caution. Rather than comparing instructed extinction with a non-instructed control group, Fuhrer and Baer instructed all participants and split them into groups based on their US expectancy ratings. Furthermore, participants who reported not expecting the US continued to show differential responding during the first block of extinction, but this responding is compared with no significant differential conditioning in participants who reported still expecting the electrotactile shock. The finding that differential responding was eliminated in all groups by the second extinction block is consistent with Wickens et al. (1963) and Notterman et al. (1952) and is unlikely to be a demonstration of resistance to instructed extinction similar to those displayed by Mandel and Bridger using a less noxious US.

Lipp, Oughton, and LeLievre (2003; Experiment 2) examined the effect of instructed extinction on electrodermal responding and a continuous measure of CS valence using a differential long delay conditioning paradigm (8s ISI – US followed CS+ immediately). During acquisition, differential first and second interval responses and differential valence evaluations were acquired between the CS+ and CS-. After instructed extinction, differential valence evaluations remained intact in both the control and the instruction group, however, no clear pattern of differential electrodermal responding was present in either the control or instruction group. Without a clear differential response in the control group, elimination of differential responding in the instruction group cannot be attributed to instructed extinction. The CS valence evaluations seemed to resist instructed extinction, however in the absence of clear instruction effects on electrodermal responding, the results of the CS valence measure should be interpreted with caution.

Sevenster, Beckers, and Kindt (2013) examined the effect of instructed extinction on electrodermal responding, blink startle, and online US expectancy throughout extinction training and after a reinstatement manipulation. In a differential long delay (7.5s ISI – US presented 0.5s before CS+ offset) conditioning design, differential electrodermal responding, blink startle modulation, and US expectancy ratings were acquired throughout acquisition training in both the control and the instruction group. Following instructed extinction, differential US expectancy ratings and entire interval electrodermal responding was intact in the control group, but eliminated in the instruction group. Differential startle modulation remained intact in both the control and the instruction groups on the first trial of extinction. Differential startle modulation was eliminated by the third extinction trial in the instructed group, while remaining intact across 11 extinction trials in the control group. Interestingly, differential US expectancy ratings re-

emerged after a subsequent reinstatement manipulation in the control group, but not the instruction group, however no other between group differences emerged after reinstatement.

Across two experiments, Luck and Lipp (2015a) examined the effect of instructed extinction using a differential long delay conditioning paradigm (6s ISI – US followed CS+ immediately), measuring electrodermal responding (Experiment 1), blink startle modulation (Experiment 2), and online CS valence (Experiment 1 and 2). In Experiment 1, differential first and second interval electrodermal responding and differential valence evaluations were acquired throughout acquisition. Following instructed extinction, differential first and second interval electrodermal responding was eliminated in the instruction group on the first extinction block (2 trials). Differential first interval responding was eliminated in controls due to an increase in responding to CS-, but differential second interval responding was still intact. In contrast, differential CS valence evaluations were not affected by instructed extinction, with intact differential valence evaluations present in both groups and no effect of instruction across extinction. In Experiment 2, differential startle modulation and differential valence evaluations were acquired in both groups. Following instructed extinction, differential startle was eliminated in the instruction group during the first block, but still intact in the control group. Differential valence ratings remained intact in both the control and the instruction group during the first block and valence evaluations did not differ between groups throughout extinction. In a third experiment, participants were asked to predict the outcome of an instructed extinction experiment after reading a detailed description of the procedure. Participants predicted that physiological responding would not change and CS+ valence would become more pleasant after instructed extinction. As these predictions were in the opposite direction to that observed in the

experiments, the authors argue that the CS valence results are unlikely to reflect demand characteristics.

Luck and Lipp (2015b) examined whether the removal of the US electrode could be responsible for mediating instructed extinction effects by comparing an instruction (electrode attached) group, an instruction (electrode removed) group, and a non-instructed control group. Using a differential long delay conditioning paradigm (6s ISI – US followed CS+ immediately), electrodermal responding and online CS valence was assessed. Throughout acquisition, differential first and second interval electrodermal responding and differential valence evaluations were acquired in all groups. Following instructed extinction, differential second interval electrodermal responding was intact in the control group, whereas differential first and second interval responding was eliminated in both instruction groups. Similar to Luck and Lipp (2015a) differential first interval responding was eliminated in the control group due to increased responding to the CS-. Differential valence evaluations were not affected by instructed extinction, with intact differential valence present in all three groups at the beginning of extinction and no interaction with group throughout extinction training.

Summary

The research examining instructed extinction of fear conditioned to non-fear relevant stimuli has confirmed that it is effective at reducing conditioned fear across a number of different conditioning designs, this reduction, however, is not always evident on the first extinction trial. Fear as indicated by electrodermal responding, heart rate, blink startle responding, and finger withdrawal seems to be subject to instructed extinction. If self-reports of conditional stimulus valence are measured, however, instructed extinction has been consistently shown not to have an effect. A number of potential moderators of the intervention have been explored, but many of

these investigations have not yielded consistent results. Silverman (1960) suggests that instructed extinction may not affect fear after conditioning with a long trace interval, but Notterman et al. (1952) used a long trace interval and found a reduction of conditional responding. Lindley and Moyer (1961) found some evidence that instructed extinction effects were stronger after minimal acquisition training, but Grings and Lockhart (1963) found no evidence that the number of acquisition trials moderated instructed extinction effects. Bridger and Mandel (1965) report that instructed extinction effects do not differ after partial or continuous reinforcement training. Wickens et al. (1963) and Grings and Lockhart (1963) directly examined instructed extinction effects after acquisition training with different US intensities, and both report that US intensity did not moderate the effects. When a very intense US was used, however, Bridger and Mandel (1965) and Mandel and Bridger (1967) report that instructed extinction did not reduce conditional responding. Despite these minor inconsistencies, instructed extinction has been shown to be a robust and reliable manipulation that will facilitate extinction and in some cases eliminate conditional responding on the very first extinction trial unless fear is indexed by CS valence evaluations and possibly after fear conditioning with a very intense US.

Instructed Extinction with Fear Relevant Conditional Stimuli

In 1970, Seligman proposed that stimuli which posed a survival threat to ancestral humans were evolutionary prepared to associate with aversive events. Prepared associations were said to be rapidly acquired, resistant to extinction, and resistant to cognitive influence (for a review see: Mallan, Lipp, & Cochrane, 2013). After this proposal, the instructed extinction manipulation became a way of assessing the proposed resistance to cognitive influence. To date, the instructed extinction manipulation has been used to examine three classes of fear relevant stimuli – phylogenetic animal fear relevant stimuli (snakes and spiders), social fear relevant

stimuli (angry faces and other race faces), and ontogenetic (modern) fear relevant stimuli (guns). In this section we will review the instructed extinction studies which used these three classes of stimuli. Additional details of the experiments can be found in Table 2 (snakes and spiders) and Table 3 (social and ontogenetic stimuli).

Phylogenetic Animal Fear Relevant Conditional Stimuli (Snakes and Spiders)

Öhman, Erixon, and Löfberg (1975) examined whether fear conditioned to fear relevant animals (snakes) would resist instructed extinction in comparison with fear conditioned to fear irrelevant pictures (houses and faces). A single cue long delay conditioning design (8s ISI – US followed CS immediately) was used, measuring electrodermal responding and manipulating fear relevance between-groups. Conditioning was present in both first and second interval electrodermal responding by the end of acquisition in all groups. After instructed extinction, second interval responding extinguished rapidly in all groups, but conditioning effects were still present in the first interval response of both fear relevant groups (instruction and control). Conditioning effects, however, were absent in both fear-irrelevant groups (instruction and control) and therefore resistance to instruction in the fear-irrelevant instruction group cannot be compared against a baseline instruction control group.

Hugdahl and Öhman (1977) replicated this finding using a differential long delay (ISI 8s – US on CS+ offset) conditioning design. Fear was conditioned to pictures of snakes and spiders (fear relevant group) and pictures of circles and triangles (fear irrelevant group). During acquisition, differential first and second interval electrodermal responding was acquired in all groups. Following instructed extinction, differential first interval responding was eliminated in the instructed fear irrelevant group, but still present in the non-instructed fear irrelevant group. In contrast, differential first interval responding remained intact in both fear relevant groups

throughout extinction. Intact differential second interval responding was present in both fear irrelevant groups throughout extinction, but in neither fear relevant group.

Hugdahl (1978) examined whether fear conditioned to pictures of snakes and spiders would resist instructed extinction after a threat of shock acquisition phase. A differential long delay conditioning design (8s ISI – US followed CS+ immediately) was used, comparing fear conditioned to images of snakes and spiders (fear relevant) with fear conditioned to images of circles and triangles (fear irrelevant). One group of participants received CS-US pairings during acquisition (conditioning group), whereas another group were told that the CS+ image would sometimes be followed by an electrotactile shock (threat group; the US was never presented). After acquisition, all participants were informed that the US would no longer be presented and the shock electrode was removed. During acquisition, differential first and second interval responding was acquired in all groups. Regardless of the conditioning procedure used during acquisition, differential first interval responding was intact in both the conditioning and threat fear relevant groups after instructed extinction. In contrast, differential first interval responding was abolished by instructions in the fear irrelevant groups. There was a rapid decrease of differential second interval responding in the fear irrelevant groups in comparison with the fear relevant groups.

Cook, Hodes, and Lang (1986; Experiment 4) examined whether the tactile component of the shock was critical to the preparedness effects which had been observed by Öhman and his colleagues. Fear was conditioned to fear relevant (snakes and spiders) and neutral pictures with a US consisting of a loud noise and vibratory stimulus to the hand. Little detail about the experiment or analysis is included in the paper, but the authors report no differential effect of instructed extinction on fear relevant and fear irrelevant groups. Cook, Hodes, and Lang (1986;

Experiment 6) used a differential long delay conditioning design (8s ISI – US followed CS+ immediately to compare the effects of instructed extinction on conditional electrodermal and heart rate responding to fear relevant (snakes and spiders) and fear irrelevant (flowers and mushrooms) stimuli after conditioning with an electrotactile shock US or a loud noise US. Differential first interval electrodermal responding developed during acquisition in both the fear relevant and fear irrelevant groups. Instructed extinction reduced first interval electrodermal responding in all instruction groups and differential responding remained only in the no instruction fear relevant shock group. A similar pattern of results was obtained with heart rate responding confirming that in this experiment fear conditioned to snakes and spiders did not resist instructed extinction.

Soares and Öhman (1993) examined the effects of instructed extinction on electrodermal conditional responding to fear relevant (snakes and spiders) or fear irrelevant (flowers and mushrooms) stimuli that were presented either backwardly masked or unmasked during extinction. Participants were conditioned in a differential short delay conditioning design (0.5s ISI – US followed CS+ immediately) and assigned to one of four groups – extinction with masked fear relevant stimuli, masked fear irrelevant stimuli, non-masked fear relevant stimuli, or non-masked fear irrelevant stimuli. Half of the participants within each of these groups were given extinction instructions, whereas, the remaining half were not informed. During acquisition, responding to CS+ was larger than responding to CS- in all groups. When extinction was performed without the mask and without instruction differential responding remained for both fear relevant and fear irrelevant stimuli. Instruction extinction, however, eliminated differential responding to neutral stimuli, but left differential responding to both masked and unmasked fear-relevant stimuli intact (but reduced in magnitude).

Lipp and Edwards (2002) aimed to replicate reports that images of snakes and spiders resist instructed extinction and to assess whether instructed extinction influenced CS valence evaluations. Using a differential long delay conditioning procedure (8s ISI – US presented at CS+ offset) participants were conditioned with fear relevant (snakes and spiders) or fear irrelevant (flowers and mushrooms) images. Participants rated the valence of the images on a 7 point Likert scale (-3 *unpleasant* to +3 *pleasant*) before and after conditioning and electrodermal responding was measured throughout the experiment. During acquisition, all groups acquired differential first and second interval responding. After instructed extinction, differential second interval responding was eliminated in the fear-irrelevant instruction group, but remained in the fear-irrelevant control group. Differential second interval responding remained in both the instructed and control fear relevant groups. There was no evidence for a differential effect of instructed extinction on the first interval electrodermal responding, however similar to Luck and Lipp (2015a; 2015b) this was likely due to an increase in responding to the CS- in the fear irrelevant control group. Evidence for conditioning was obtained in the CS valence measure but this did not interact with the instructional manipulation. This finding could suggest that instructed extinction did not affect the CS valence evaluations, but should be interpreted with care due to the limitations involved in using a post extinction assessment of valence.

Luck and Lipp (under review; Experiment 1) aimed to replicate resistance to instructed extinction for fear conditioned to images of snakes and spiders using a within-participants design. The between-participants design has been criticised as the repeated exposure to fear eliciting stimuli in the fear relevant group could lead to between group differences in state anxiety which could affect conditioning (Mertens, Raes, & De Houwer, 2016). Using a differential long delay conditioning design (6s ISI – US presented at CS+ offset), participants

viewed images of two fear relevant (snake and spider) and two fear irrelevant (bird and fish) animals. One picture from each fear relevance category was used as CS+ and the other as CS-. Differential first and second interval responding was acquired to both fear relevant and fear irrelevant images throughout acquisition. After instructed extinction, differential second, but not first, interval responding remained intact to fear relevant images on the first extinction trial, whereas differential first and second interval responding to fear irrelevant images was eliminated.

Social and Ontogenetic Fear Relevant Stimuli

Mallan, Sax, and Lipp (2009) assessed the influence of instructed extinction on blink startle modulation and first interval electrodermal responding after conditioning with racial in-group or out-group faces. A long delay differential conditioning design (6s ISI – US presented at CS+ offset) was used and Chinese male faces were used as the racial outgroup within a group of Caucasian participants (most appropriate racial in and out-groups in Australia). During acquisition, differential startle modulation and differential electrodermal responding was acquired in all groups. Following instructed extinction, the control group conditioned with out-group faces continued to show differential electrodermal and startle responding, but differential responding was extinguished in instructed participants conditioned with out-group faces. Differential responding was not present in participants conditioned with in-group faces throughout extinction, regardless of instruction group.

As part of a larger study, Olsson and Phelps (2004) examined the effect of instructed extinction on fear conditioned to angry faces after an instructed acquisition phase. Participants were informed that the CS+ would be paired with the electrotactile shock (US was never actually presented) and that the CS- would be presented alone. Differential responding was not present

during acquisition, however the acquisition analyses were focused on a subset of masked trials and it is unclear whether differential responding was present on the unmasked trials. After instructed extinction, differential responding was present between CS+ and CS- and was maintained during extinction. This finding suggests that fear conditioned to angry faces may resist instructed extinction, but this conclusion should be interpreted with care as differential responding was not present during acquisition and the experiment was not designed to assess instruction effects as it was a small part of a larger study. Rowles, Lipp, and Mallan (2012) examined the effect of instructed extinction on fear conditioned to angry faces directly using a differential long delay conditioning design (6s ISI – US presented at CS+ offset). During acquisition, one group of participants was conditioned with images of angry faces and another with images of happy faces. Both groups acquired differential first interval electrodermal responding, but after instructed extinction only the angry control group showed differential responding, suggesting that fear conditioned to angry faces does not resist instructed extinction. A pre-post measure of CS valence showed evidence of conditioning but this did not interact with the instructional manipulation.

Luck and Lipp (under review; Experiment 2) used a within-participants instructional design to examine whether fear conditioned to images of pointed guns would resist instructed extinction. Using a within participants differential long delay conditioning design (6s ISI – US presented at CS+ offset), participants viewed images of pointed guns (fear relevant) and pointed hairdryers (fear irrelevant). Throughout acquisition, differential first and second interval electrodermal responding was evident to images of guns and hairdryers, however following instructed extinction, differential first and second interval electrodermal responding to both sets of images was eliminated.

Summary

The instructed extinction manipulation has been used in a number of studies to assess whether, as suggested by preparedness theory, fear conditioned to a range of fear relevant CSs is encapsulated from cognition. There is substantial evidence that fear conditioned to images of snakes and spiders is not sensitive to instructed extinction. Of the eight studies designed to investigate this, five (Öhman et al., 1975; Hugdahl & Öhman, 1977; Hugdahl, 1978; Soares & Öhman, 1993; Lipp & Edwards, 2002; and Luck and Lipp, under review) have reported that fear conditioned to snakes and spiders resists instructed extinction. There has been little evidence, however, that fear conditioned to other classes of fear relevant stimuli resists instructed extinction. Fear conditioned to other race faces (Mallan, Sax, & Lipp, 2009), angry faces (Rowles, Lipp, & Mallan, 2012), and pointed guns (Luck & Lipp, under review) was reduced after instructed extinction.

Integration, Clinical Applications, and Future Directions

It is clear that instructed extinction has a long and rich history within human fear conditioning experiments. Instructed extinction experiments have used short and long CS durations, single cue and differential conditioning paradigms, different reinforcement rates and amounts, and a number of different conditional and unconditional stimuli. Despite this variation, the pattern of instructed extinction effects is remarkably consistent – instructed extinction reduces conditional fear as indexed by electrodermal responding, startle modulation, heart rate, conditioned finger withdrawal responding and US expectancy ratings. This effect is not always present on the first trial of extinction, but with only a few exceptions, instructed extinction does facilitate the extinction of conditioned fear.

The majority of studies have not assessed the effect of instructed extinction on the first trial of extinction, and in those studies which have the results are mixed. Some authors report that conditional responding is eliminated prior to explicit extinction training, but others report that instructed extinction only facilitates extinction learning. As instructed extinction has been shown to eliminate conditional responding on the first extinction trial in a number of studies, it is possible that factors which vary across studies, such as the control of participant beliefs, could be influencing the results. Participants' belief in the instructions is a very powerful factor and inclusion of participants who are sceptical about the validity of the instructions could mask instruction effects on the first trial of extinction (Luck & Lipp, 2015b; Mandel & Bridger, 1973).

Across the literature there have been three notable exceptions to the general pattern of instructed extinction results – instructed extinction does not affect conditional stimulus valence; fear conditioned to snakes and spiders survives instructed extinction; and fear conditioned with a very painful electrotactile shock may resist instructed extinction. One potential explanation of these exceptions may be that emotional conditioning, prepared stimuli, and intensely aversive stimuli activate a subcortical fear processing system which is more resistant to cognitive influence (Debiec & LeDoux, 2004; Öhman, 2005). More research is needed, however, to examine whether there are more parsimonious explanations which could also account for these exceptions.

These 'exceptions' observed in the laboratory may have implications for clinical practice, however, there are limitations to the extent to which fear conditioned in the laboratory with an unpleasant US compares to the experiences of an individual suffering from, for instance, post-traumatic stress disorder. Nevertheless, differences in response to instruction observed across experiments may also manifest in clinical practice. The observation that fear conditioned with a

very painful shock resists instructed extinction may suggest that fear responses seen in the clinic which have been acquired based on intensely aversive real live experiences may be less responsive to cognitive intervention. Similarly, if fear conditioned to snakes and spiders, but not other animals, resists instruction in the laboratory, snake and spider phobias may require different approaches than those used for other small animal phobias. If there is a dissociation between the subjective dislike of feared situations and events and physiological responding after instructed extinction, then similar dissociations may be observed after successful treatment. Persisting negative valence predicts higher reinstatement rates after fear extinction (Dirkx, Hermans, Vansteenwegen, & Baeyens, 2004; Hermans et al., 2005; Zbozinek, Hermans, Prenoveau, Liao, & Craske, 2015) and manipulations which reduce negative CS+ valence have been shown to reduce fear reinstatement (Zbozinek, Holmes, & Craske, 2015)

Instructed extinction is proposed as a laboratory analogue for cognitive interventions, but falls short of capturing the complexity of cognitive interventions used in the clinical setting. Instructed extinction completely breaks the association between the feared stimulus and the aversive event, whereas, cognitive therapy is used to bring the probability of negative outcomes more in line with reality. The robust decreases in physiological responding observed after instructed extinction may occur because of the certainty involved in the manipulation. Future research should examine the use of instructional manipulations which weaken the CS-US contingency, without breaking it completely. As a probability based cognitive manipulation, instructed extinction does not capture a number of other aspects often targeted throughout cognitive therapy, such as reappraising the cost of the aversive event occurring and the client's ability to handle an aversive event if it was to occur. Negative valence, fear of snakes and spiders, and fears acquired based on very aversive events may still respond to these other aspects

of cognitive therapy. In support of this idea negative CS+ valence can be removed with a cognitive intervention specifically targeting CS valence, rather than CS-US contingency (Luck & Lipp, under revision). More research is required to disentangle the components involved in cognitive therapy, to examine the reliability of instructed extinction as an analogue for cognitive interventions, and to examine whether different types of cognitive interventions would be more effective at targeting negative valence and more robust fear responses.

Sevenster et al's. (2013) is the only study to date to have assessed the effects of instructed extinction on the return of fear directly. In this study, instructed extinction did not influence the reinstatement of differential electrodermal responding or startle modulation but did reduce the return of differential US expectancy ratings. This initial finding is promising, but more follow-up research is needed to assess the effects of instructed extinction on the return of fear using renewal and spontaneous recovery procedures. Instructed extinction research in the laboratory has provided researchers with a number of interesting 'exceptions' which do require further study, but their implications should also be examined in clinical settings. Are cognitive interventions less effective for treatment of snake and spider phobias? Are they less effective when fear has been acquired in an intensely traumatic or negative situation? Is it possible to change the valence of the feared stimulus in the clinical setting and does this reduce relapse? Instructed extinction research has come a long way since the first study was published in 1937, and now seems the time to translate some of its findings and implications to clinically based applied research.

Table 1. Instructed Extinction Research Using Non-Fear Relevant Conditional Stimuli

Electrodermal Responding										
Study	**Conditioning Design**	**CS**	**US**	**Conditioning and ISI**	**Acquisition**	**Extinction**	**Instruction Comparison**	**First Trial Effects**	**Facilitation?**	**Electrode Removal? (Instructed Groups)**
Cook & Harris (1937)	Single cue (no control)	3s light	Shock (duration not reported)	Delay – 3s (US on CS offset)	30 CS-US pairings (100% reinforcement)	Not specified	Between groups – control vs. instruction	Not assessed	Yes	Not specified
Silverman (1960)	Single cue (unpaired control)	2s tone	6s shock	Trace – 2.5s and 8s ISI (0.5s and 6s interval between CS and US)	10 CS-US pairings (100% reinforcement)	15 CS alone trials	Between groups – instruction vs. control (in each ISI condition)	Not assessed	2.5s ISI: Yes 8s ISI: No Unpaired: Yes	Not specified
Wickens, Allen & Hill (1963)	Single cue (unpaired control)	0.5s tone	0.1s shock	Delay – 0.5s ISI (US on CS offset)	10 CS-US trials (Strong US group: CS paired with a strong shock (with 10 weak shocks interspersed between trials); Weak US group: CS paired with a weak shock (with 10 strong shocks interspersed between trials) Control: 10 strong and 10 weak shocks (unpaired with the CS) (100% reinforcement)	5 CS alone trials	Between groups – instruction vs. control (in each US intensity group)	No	Yes	Yes
Grings and Lockhart (1963)	Single cue (no control)	5s pictures (shapes)	Shock (duration not reported)	Delay – 5s ISI (US on CS offset)	Minimum reinforcement: 9 (3 of each CS) (100% reinforcement) Extended reinforcement: 36 (12 of each CS) (100% reinforcement)	3 CS alone trials (1 of each CS)	Between groups – instruction vs. control (in each reinforcement condition)	Yes	Not assessed (only one of each CS trial)	Not specified
Bridger & Mandel (1964)	Differential conditioning	6s lights	0.5s shock (very painful)	Delay – 6s ISI (US delivered 1s before CS offset)	Shock group: 20 CS+ – US pairings (100% reinforcement). 20 CS- alone.	10 CS+ and 10 CS- trials (unreinforced)	Between groups – instruction vs. control (in each of the shock and threat groups)	No	No	Yes

					Threat group: 20 CS+ alone trials (threatened). 20 CS- alone trials.					
Bridger & Mandel (1965)	Differential conditioning	0.5s lights	0.5s shock (very painful)	Delay – 0.5s ISI (US on CS offset)	20 CS+ (5/20 reinforced in partial reinforcement group and 20/20 reinforced in continuous reinforcement group). 20 CS- alone trials.	30 CS+ and 30 CS- trials (unreinforced)	Between groups – instruction vs. control (in each of the partial and continuous reinforcement groups)	Significant reduction (but continued differential responding)	Yes	Yes
Mandel & Bridger (1967)	Differential conditioning	Short ISI group: 0.5s lights. Long ISI group: 5s lights	0.5s shock (very painful)	Delay – 0.5s or 5s ISI (US on CS offset)	25 CS+ trials (15/25 reinforced). 25 CS- alone trials	10 CS+ and 10 CS- trials (unreinforced)	Between groups – instruction vs. control (in each acquisition group)	No	Not possible to assess	Yes
Fuhrer & Baer (1980)	Differential conditioning	Short ISI group: 0.5s tones. Long ISI group: 8s tones	0.25s shock	Delay – 0.5s or 8s ISI (US on CS offset)	30 CS+ presentations (18 reinforced). 30 CS- presentations	10 CS+ and 10 CS- trials (unreinforced)	No control – all participants receive instructed extinction manipulation. Participants later split based on US expectancy scores.	No	Yes	Yes
Lipp, Oughton & LeLivre (2003)	Differential conditioning	8s pictures of vowels	.5s shock	Delay – 8s ISI (US on CS offset)	10 CS+ presentations (100% reinforcement). 10 CS- alone presentations.	16 CS+ and 16 CS- presentations (unreinforced)	Between groups – Control vs. instruction	Not possible to assess	Not possible to assess	Yes
Sevenster, Beckers, & Kindt (2012)	Differential conditioning	8s pictures of shapes	2ms shock	Delay – 7.5s ISI (US presented 7.5s into 8s CS)	6 CS+ presentations (4 reinforced). 6 CS- alone presentations. (Acquisition on Day 1)	16 presentations of CS+ and CS- (unreinforced) (extinction on day 2)	Between groups – Control vs. instruction	Yes	Yes	No
Luck and Lipp (2015a; Experiment 1)	Differential conditioning	6s pictures of neutral faces	0.2s shock	Delay – 6s ISI (US on CS offset)	8 CS+ presentations (100% reinforcement). 8 CS- alone trials.	8 CS+ and 8 CS- presentations (unreinforced)	Between groups – Control vs. instruction	Yes [1]	Yes	Yes
Luck and Lipp (2015b)	Differential conditioning	6s pictures of neutral faces	0.2s shock	Delay – 6s ISI (US on CS offset)	8 CS+ presentations (100% reinforcement). 8 CS- alone trials.	8 CS+ and 8 CS- presentations (unreinforced)	Between groups – Control vs. instruction (electrode attached)	Yes	Yes	Electrode attached: No Electrode removed: Yes

							vs. instruction (electrode removed)			
Heart Rate										
Notterman, Schoenfeld & Bersh (1952)	Single cue (no control)	1s tone	6s shock	Trace –7s ISI (6s interval between CS and US)	18 CS presentations (11 reinforced – 61%)	11 CS (unreinforced) First extinction trial excluded	Between groups – Control vs. instruction	No	Yes	Not specified
Blink Startle Responding										
Sevenster, Beckers, & Kindt (2012)	Differential conditioning	8s pictures of shapes	2ms shock	Delay – 8s ISI (US presented 7.5s into 8s CS)	6 CS+ presentations (4 reinforced). 6 CS- alone presentations. (acquisition on day 1)	16 presentations of CS+ and CS- (unreinforced) (extinction on day 2)	Between groups – Control vs. instruction	No	Yes	No
Luck and Lipp (2015a; Experiment 2)	Differential conditioning	6s pictures of neutral faces	0.2s shock	Delay – 6s ISI (US on CS offset)	8 CS+ presentations (100% reinforcement). 8 CS- alone trials.	12 CS+ and 12 CS- presentations (unreinforced)	Between groups – Control vs. instruction	Not possible to assess [2]	Yes	Yes
Finger Withdrawal										
Lindley & Moyer (1961)	Single cue (no control)	0.5s tone	0.2s shock	Trace – 1s ISI (.5s interval between CS and US)	Minimum reinforcement: until participant reached criterion of 4 conditioned responses in 5 consecutive trials (average 21 pairings). Extended reinforcement: 20 additional conditioning trials after reaching criterion	25 CS trials (unreinforced)	Between groups – Control (no instructions/no pause) vs. Control (interrupted but no information given) vs. Instructed (informed to let the finger move automatically) vs. Instructed (informed to suppress finger movement)	Not assessed	Yes	Not specified (unlikely as shock embedded within experimental set-up)
US Expectancy										
Sevenster, Beckers, & Kindt (2012)	Differential conditioning	8s pictures of shapes	2ms shock	Delay – 8s ISI (US presented 7.5s into 8s CS)	6 CS+ presentations (4 reinforced). 6 CS- alone presentations. (acquisition on day 1)	16 CS+ 16 CS- (unreinforced) (extinction on day 2)	Between groups – Control vs. instruction	Yes	Yes	No
CS Valence										
Lipp, Oughton &	Differential conditioning	8s pictures of vowels	.5s shock	Delay – 8s ISI (US on CS offset)	10 CS+ presentations (100% reinforcement).	16 CS+ 16 CS- trials (unreinforced)	Between groups – Control vs. instruction	No	No	Yes

LeLivre (2003)					10 CS- alone presentations.					
Luck and Lipp (2015a; Experiment 1)	Differential conditioning	6s pictures of neutral faces	0.2s shock	Delay – 6s ISI (US on CS offset)	8 CS+ presentations (100% reinforcement). 8 CS- alone trials.	8 CS+ 8 CS- trials (unreinforced)	Between groups – Control vs. instruction	No	No	Yes
Luck and Lipp (2015a; Experiment 2)	Differential conditioning	6s pictures of neutral faces	0.2s shock	Delay – 6s ISI (US on CS offset)	8 CS+ presentations (100% reinforcement). 8 CS- alone trials.	12 CS+ and 12 CS- (unreinforced)	Between groups – Control vs. instruction	No	No	Yes
Luck and Lipp (2015b)	Differential conditioning	6s pictures of neutral faces	0.2s shock	Delay – 6s ISI (US on CS offset)	8 CS+ presentations (100% reinforcement). 8 CS- alone trials.	8 CS+ 8 CS- trials (unreinforced)	Between groups – Control vs. instruction (electrode attached) vs. instruction (electrode removed)	No	No	Electrode attached: No Electrode removed: Yes

Notes:[1] Instruction effects were analysed in this study based on blocks – a reanalysis of the electrodermal responding data based on trials revealed that differential responding was not present on the first trial. [2] The first startle probe was in the second trial of extinction.

Table 2. Instructed Extinction Research using Phylogenetic Animal Fear Relevant Stimuli

Electrodermal Responding										
Study	**Conditioning Design**	**CS**	**US**	**Conditioning and ISI**	**Acquisition**	**Extinction**	**Instruction Comparison**	**First Trial Effects**	**Facilitation?**	**Electrode Removal?**
Öhman, Erixon, and Löfberg (1975)	Single cue (no pseudo-conditioning control)	8s slides of snakes, houses, and faces	50ms shock	Delay – 8s ISI (US on CS offset)	10 presentations of snakes, houses, and faces (snakes paired with US for one group, houses paired with US for another, and faces paired with US for the third group)	10 snakes 10 houses 10 faces (unreinforced)	Between groups – instructed vs. control	Not assessed	Fear irrelevant stimuli: Not possible to assess. Fear relevant stimuli: No	Yes
Hugdahl & Öhman (1977)	Differential conditioning	8s slides of a snake and spider (fear relevant) or a triangle and a circle (fear irrelevant)	Shock (duration not reported)	Delay – 8s ISI (US on CS offset)	10 CS+ (100% reinforcement) 10 CS- alone	14 CS+ 14 CS- (unreinforced)	Between groups – instructed vs. control (in each of the fear relevant and fear irrelevant groups)	Not assessed	Fear irrelevant stimuli: Yes Fear relevant stimuli: No	Yes
Hugdahl (1978)	Differential conditioning	8s pictures of a snake and spider (fear relevant) or a triangle and a circle (fear irrelevant)	Shock (duration not reported)	Delay – 8s ISI (US on CS offset)	12 CS+ (100% reinforcement) 12 CS- alone	20 CS+ 20 CS- (unreinforced)	All participants received instructed extinction manipulation	Not assessed	Fear irrelevant stimuli: Yes Fear relevant stimuli: No	Yes
Cook, Lang, & Hodes (1986-Experiment 4)	Differential conditioning	8s slides of snakes and spiders or neutral stimuli	Loud noise and vibrotactile sensation to arm (duration not reported)	Delay – 8s ISI (US on CS offset)	Not specified	Not specified	Between groups – instructed vs. control (in each of the fear relevant and fear irrelevant groups)	No	No	Yes

Cook, Lang, & Hodes (1986-Experiment 6)	Differential conditioning	8s slides of snakes and spiders (fear relevant) or flowers and mushrooms (fear irrelevant)	0.5s shock or 0.5s loud noise (between participants)	Delay – 8s ISI (US on CS offset)	8 CS+ (100% reinforcement) 8 CS- presented alone	20 CS+ 20 CS- trials (unreinforced)	Between groups – instructed vs. control (in each of the fear relevant shock and noise and fear irrelevant shock and noise groups)	No	No	Yes
Soares & Öhman (1993)	Differential conditioning	0.5s, 30ms, or 0.13s slides of snakes and spiders (fear relevant) or flowers and mushrooms (fear irrelevant)	0.5s shock	Acquisition: Delay – 0.5s ISI (US on CS offset) Extinction: Masked group: CS presented for 30ms followed by masking stimulus for 0.1s Non-masked group: CS presented for 0.13s	12 CS+ (10 reinforced) 12 CS- (unreinforced)	16 CS+ and 16 CS- (unreinforced)	Between groups – instructed vs. control (in each of the masking conditions)	Not assessed	Fear Relevant Stimuli: No Fear irrelevant stimuli: Yes	Yes
Lipp & Edwards (2002)	Differential conditioning	8s images of snakes and spiders (fear relevant) or flowers and mushrooms (fear irrelevant)	0.2s shock	Delay – 8s ISI (US on CS offset)	10 CS+ (100% reinforcement). 10 CS- (unreinforced)	8 CS+ 8 CS- (unreinforced)	Between groups – instructed vs. control (in each of the fear relevance categories)	Not assessed	Fear Relevant: No Fear irrelevant: Yes	Yes
Luck and Lipp (under review; Experiment 1)	Differential conditioning	6s images of fear relevant (snakes/spiders) and fear irrelevant (birds/fish)	0.2s shock and loud noise combination	Delay – 6s ISI (US on CS offset)	6 CS+ (100% reinforcement). 6 CS- alone (for both fear relevant and fear irrelevant stimuli)	6 CS+ and 6 CS- unreinforced trials (for both fear relevant and fear irrelevant stimuli)	Within-groups – all participants received extinction instructions.	Fear Relevance: No Fear irrelevant: Yes	Fear Relevance: No Fear irrelevant: Yes	Yes
Heart Rate										

Cook, Lang, & Hodes (1986-Experiment 6)	Differential conditioning	8s slides of snakes and spiders (fear relevant) or flowers and mushrooms (fear irrelevant)	0.5s shock or 0.5s loud noise (US varied between participants)	Delay – 8s ISI (US on CS offset)	8 CS+ (100% reinforcement) 8 CS- presented alone	20 CS+ and 20 CS- trials (unreinforced)	Between groups – instructed vs. control (in each of the fear relevant shock and noise and fear irrelevant shock and noise groups)	No	No	Yes

Table 3. Instructed Extinction Research using Social and Ontogenetic Fear Relevant Stimuli

Electrodermal Responding										
Study	**Conditioning Design**	**CS**	**US**	**Conditioning and ISI**	**Acquisition**	**Extinction**	**Instruction Comparison**	**First Trial Effects**	**Facilitation?**	**Electrode Removal?**
Olsson & Phelps (2004)	Differential conditioning (threat of shock acquisition)	6s pictures of angry faces	Threat of shock (no actual presentations)	6s CS duration – no US presentations (in the masked group on masked trials the CS was presented for 33ms followed by the mask)	12 CS+ (unreinforced) 12 CS- (unreinforced)	10 CS+ and 10 CS- trials (unreinforced)	All participants in this part of the experiment received extinction instructions	No	No	Not Specified
Mallan, Sax, & Lipp (2009)	Differential conditioning	6s pictures of Chinese faces or Caucasian faces (all males)	0.4s shock	Delay – 6s ISI (US on CS offset)	8 CS+ (100% reinforcement) 8 CS- presented alone	12 CS+ and 12 CS- trials (unreinforced)	Between groups – instructed vs. control (in each of the fear relevance categories)	Not assessed	Yes	Yes
Rowles, Lipp, & Mallan (2012)	Differential conditioning	6s pictures of happy or angry faces (males)	0.2s shock	Delay – 6s ISI – (US on CS offset)	8 CS+ (100% reinforcement) 8 CS- presented alone	10 CS+ and 10 CS- trials (unreinforced)	Between groups – instructed vs. control (in each of the fear relevance categories)	Not assessed	Yes	Yes
Luck and Lipp (under review; Experiment 2)	Differential conditioning	6s pictures of fear relevant (guns) and fear irrelevant (hairdryers)	0.2s shock and loud noise combination	Delay – 6s ISI – (US on CS offset	6 CS+ (100% reinforcement). 6 CS- alone (for both fear relevant and fear irrelevant stimuli)	6 CS+ and 6 CS- alone trials (for both fear relevant and fear irrelevant stimuli)	Within-groups – all participants received extinction instructions.	Fear Relevance: Yes Fear irrelevant: Yes	Fear Relevance: Yes Fear irrelevant: Yes	Yes

Blink Startle										
Mallan, Sax, & Lipp (2009)	Differential conditioning	6s pictures of Chinese faces or Caucasian faces (males)	0.4s shock	Delay – 6s ISI – (US on CS offset	8 CS+ (100% reinforcement) 8 CS- presented alone	12 CS+ and 12 CS- trials (unreinforced)	Between groups – instructed vs. control (in each of the fear relevance categories)	Not assessed	Yes	Yes

References

Bisson, J., & Andrew, M. (2007). Psychological treatment of post-traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews*, 2007 (3), doi: 10.1002/14651858.CD003388.pub3.

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

Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L.G. Tassinary & G.G. Bernston (Eds.), (2000). Handbook of Psychophysiology (pp. 602-642). New York: Cambridge University Press.

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. doi: http://dx.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. doi: 10.1037/h0021764

Craske, M. G. (1999). Anxiety disorder: Psychological approaches to theory and treatment, Boulder, CO: Westview Press.

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

Cook, S. W., & Harris, R. E. (1937). The verbal conditioning of the galvanic skin reflex. *Journal of Experimental Psychology, 21*, 202-210. doi: 10.1037/h0063197

Cook, E. W., Hodes, R. L., & Lang, P. J. (1986). Preparedness and phobia: Effects of stimulus content on human visceral conditioning. *Journal of Abnormal Psychology, 95*(3), 195-207. doi: 10.1037/0021-843X.95.3.195

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology, 33*, 103-111. doi: 10.1111/j.1469-8986.1996.tb02114.x

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L.G. Tassinary & G.G. Bernston (Eds.), (2007). Handbook of Psychophysiology (pp. 159-181). Cambridge: Cambridge University Press.

Debiec, J., & LeDoux, J. (2004). Fear and the Brain. *Social Research, 71*(4), 807-818.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*, 549-554. doi: 10.1101/lm.78004

Fuhrer, M. J., & Baer, P. E. (1980). Cognitive factors and CS-UCS interval effects in the differential conditioning and extinction of skin conductance responses. *Biological psychology, 10*(4), 283-298. doi: http://dx.doi.org/10.1016/0301-0511(80)90041-1

Grings, W. W., & Lockhart, R. A. (1963). Effects of "anxiety-lessening" instructions and differential set development on the extinction of GSR. *Journal of Experimental Psychology, 66*(3), 292-299. doi: 10.1037/h0045094

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

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

Hugdahl, K. (1978). Electrodermal conditioning to potentially phobic stimuli: Effects of instructed extinction. *Behaviour Research and Therapy, 16*(5), 315-321. doi: http://dx.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 and Memory, 3*(5), 608-618. doi: 10.1037/0278-7393.3.5.608

Kessler, R. C., Koretz, D., Merikangas, K. R., & Wang, P.S. (2004). The epidemiology of adult mental disorders. In B.L. Levin, J. Petrilia, & K.D. Hennessy (Eds.), (2004). Mental health services: A public health perspective. New York: Oxford University Press.

Lang, P. J. (1985). The cognitive psychophysiology of emotion: Fear and anxiety, Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological review, 97*(3), 377-395. doi: 10.1037/0033-295X.97.3.377

Lindley, R. H., & Moyer, K. E. (1961). Effects of instructions on the extinction of a conditioned finger-withdrawal response. *Journal of Experimental Psychology, 61*(1), 82-88. doi: 10.1037/h0047005

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

Lipp, O. V., & Edwards, M. S. (2002). Effect of instructed extinction on verbal and autonomic indices of Pavlovian learning with fear-relevant and fear-irrelevant conditional stimuli. *Journal of Psychophysiology, 16*(3), 176-186. doi: 10.1027//0269-8803.16.3.176

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*(3), 219-239. doi: http://dx.doi.org/10.1016/S0023-9690(03)00011-0

Lipp, O. V., Sheridan, J., & Siddle, D. A. T. (1994). Human blink startle during aversive and nonaversive Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 20,* 380-389. doi: 10.1037/0097-7403.20.4.380

Lipp, O. V., Siddle, D. A. T., & Dall, P. J. (2003). The effects of unconditional stimulus valence and conditioning paradigm on verbal, skeleto-motor, and autonomic indices of human Pavlovian conditioning. *Learning and Motivation, 34,* 32-51. doi:10.1016/S0023-9690(02)00507-6

Luck, C. C., & Lipp, O. V. (2015). 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. doi: http://dx.doi.org/10.1016/j.brat.2015.01.001

Luck, C. C., & Lipp, O. V. (2015). To remove or not to remove? Removal of the unconditional stimulus electrode does not mediate instructed extinction effects. *Psychophysiology, 52*(9), 1248-1256. doi: 10.1111/psyp.12452

Luck, C. C., & Lipp, O. V. (2016). When orienting and anticipation dissociate — a case for scoring electrodermal responses in multiple latency windows in studies of human fear conditioning. *International Journal of Psychophysiology, 100,* 36-43. doi: http://dx.doi.org/10.1016/j.ijpsycho.2015.12.003

Luck, C. C., & Lipp, O. V. (under revision). Conditioned negative stimulus evaluations can be reduced with cognitive interventions targeting valence (but no evidence that this reduction moderates reinstatement rates).

Luck, C. C., & Lipp, O. V. (under review). Phylogenetic, but not ontogenetic, fear relevant stimuli resist instructed extinction in a within-participants design.

Lipp, O. V., & Vaitl, D. Reaction time task as unconditional stimulus. *The Pavlovian journal of biological science, 25*(2), 77-83. doi: 10.1007/bf02964606

Mallan, K. M., Lipp, O. V., & Cochrane, B. (2013). Slithering snakes, angry men and out-group members: What and whom are we evolved to fear? *Cognition and Emotion, 27*(7), 1168-1180. doi: 10.1080/02699931.2013.778195

Mallan, K. M., Sax, J., & Lipp, O. V. (2009). Verbal instruction abolishes fear conditioned to racial out-group faces. *Journal of Experimental Social Psychology, 45*(6), 1303-1307. doi: http://dx.doi.org/10.1016/j.jesp.2009.08.001

Mandel, I. J., & Bridger, W. H. (1967). Interaction between instructions and ISI in conditioning and extinction of the GSR. *Journal of Experimental Psychology, 74*(1), 36-43. doi: 10.1037/h0024496

Mandel, I. J., & Bridger, W. H. (1973). Is there classical conditioning without cognitive expectancy? *Psychophysiology, 10*(1), 87-90. doi: 10.1111/j.1469-8986.1973.tb01088.x

Mertens, G., Raes, A. K., & De Houwer, J. (2016). Can prepared fear conditioning result from verbal instructions? *Learning and Motivation, 53*, 7-23. doi: http://dx.doi.org/10.1016/j.lmot.2015.11.001

Mowrer, O. H. (1938). Preparatory set (expectancy)—a determinant in motivation and learning. *Psychological review, 45*, 62-91. doi: 10.1037/h0060829

Notterman, J. M., Schoenfeld, W. N., & Bersh, P. J. (1952). A comparison of three extinction procedures following heart rate conditioning. *The Journal of Abnormal and Social Psychology, 47*, 674-677. doi: 10.1037/h0061624

Öhman, A. (1983). The orienting response during Pavlovian conditioning. In D. A. T. Siddle (Ed.), *Orienting and habituation: Perspectives in human research* (pp. 315-370). New York: Wiley.

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology, 30*(10), 953-958. doi: http://dx.doi.org/10.1016/j.psyneuen.2005.03.019

Öhman, A., Erixon, G., & Löfberg, I. (1975). Phobias and preparedness: Phobic versus neutral pictures as conditioned stimuli for human autonomic responses. *Journal of Abnormal Psychology, 84*(1), 41-45. doi: 10.1037/h0076255

Ougrin, D. (2011). Efficacy of exposure versus cognitive therapy in anxiety disorders: systematic review and meta-analysis. *BMC Psychiatry, 11*(1), 1-12. doi:10.1186/1471-244X-11-200

Prokasy, W.F., & Kumpfer, K.L. (1973). Classical conditioning. In W. F. Prokasy & D. C. Raskin (Eds.), *Electrodermal activity in psychological research* (pp. 157-202). San Diego: Academic Press.

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychological Science, 15*(12), 822-828.

Quinn, J. J., & Fanselow, M.S. (2006) in Craske, M. G., Hermans, D., & Vansteenwegen, D. (Eds). (2006). Fear and learning: From basic processes to clinical implications. Washington: APA Books.

Rachman, S. (1966). Studies in desensitization III: speed of generalization. Behaviour research and therapy, *4*, 7-15. doi:10.1016/0005-7967(66)90038-6

Rachman, S. (1968). *Phobias: Their nature and control.* Illinois: Thomas.

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

Rowles, M. E., Lipp, O. V., & Mallan, K. M. (2012). On the resistance to extinction of fear conditioned to angry faces. *Psychophysiology, 49*(3), 375-380. doi: 10.1111/j.1469-8986.2011.01308.x

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

Seligman, M. E. (1970). On the generality of the laws of learning. *Psychological review, 77*(5), 406-418. doi: 10.1037/h0029790

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. doi: 10.1111/j.1469-8986.2012.01450.x

Silverman, R. E. (1960). Eliminating a conditioned GSR by the reduction of experimental anxiety. *Journal of Experimental Psychology, 59*(2), 122-125. doi: 10.1037/h0045555

Soares, J. J. F., & Öhman, A. (1993). Preattentive processing, preparedness and phobias: Effects of instruction on conditioned electrodermal responses to masked and non-masked fear-relevant. *Behaviour Research and Therapy*, *31*(1), 87-95. doi: http://dx.doi.org/10.1016/0005-7967(93)90046-W

Vervliet, B., Craske, M., & Hermans, D. (2013). Fear extinction and relapse: State of the Art. *Annual Review of Clinical Psychology, 9*, 215-48. doi:10.1146/annurev-clinpsy-050212-185542

Wickens, D. D., Allen, C. K., & Hill, F. A. (1963). Effects of instruction on extinction of the conditioned GSR. *Journal of Experimental Psychology, 66*(3), 235-240. doi: 10.1037/h0048932

Zbozinek, T. D., Hermans, D., Prenoveau, J. M., Liao, B., & Craske, M. G. (2015). Post-extinction conditional stimulus valence predicts reinstatement fear: Relevance for long-term outcomes of exposure therapy. *Cognition and Emotion, 29*(4), 654-667. . doi: 0.1080/02699931.2014.930421

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