Behaviour Research and Therapy 112 (2019) 42–62

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

The need for standards in the design of differential fear conditioning and extinction experiments in youth: A systematic review and recommendations for research on anxiety

Katherine M. Ryan*, Melanie J. Zimmer-Gembeck, David L. Neumann, Allison M. Waters**

School of Applied Psychology, Griffith University, Australia

ARTICLE INFO

Keywords:
Fear conditioning
Fear extinction
Anxiety
Youth

ABSTRACT

Fear extinction studies in youth have yielded mixed results due to developmental processes and variations in design, methodology and dependent measures. This systematic review focused on studies with healthy youth between 2 and 17 years of age to identify experimental parameters of studies documenting extinction effects. Thirty-five studies met inclusion criteria and the following themes emerged (a) some studies employed parameters and task demands that are complex and require active participant involvement whereas others involved simple stimulus configurations and passive participant involvement, and (b) variation exists among dependent measures in units of measurement, timing and type of measurement. The review identified that studies using geometric shape conditioned stimuli (CS) paired with a tone unconditioned stimulus (US) (e.g., metal scraping on slate), as well as face CSs with a scream US produced the most reliable extinction effects, although the latter combination may be associated with higher drop-out than shape CSs and a tone US. The most commonly used and effective dependent measures for revealing extinction effects were skin conductance responses (SCR) and subjective ratings (SR) of CS valence, fearfulness and arousal. Fear potentiated startle (FPS) blink reflexes were also an effective but less commonly used measure. It is recommended that future studies use shape CSs and the metal scraping on slate US in studies involving children and either shape CSs and the metal scraping on slate US or face CSs paired with a scream US with adolescents. It is also recommended that US expectancy ratings and CS evaluations are assessed trial-by-trial and between-phase, and that startle-eliciting stimuli to measure startle blink reflexes are delivered on every second trial per CS so that SCR and FPS can be examined. However, further research is required to determine whether increased participant involvement due to providing trial-by-trial and between-phase ratings of the CSs and US differentially influences responding, particularly in children relative to adolescents and adults.

1. Introduction

Exposure-based cognitive-behavioural therapy (CBT) is an efficacious and widely used psychological treatment for anxiety disorders (Higa-McMillan, Francis, Rith-Najaran & Chorpita, 2016), in which gradual and repeated exposure to the feared object or situation is a key component (Kendall et al., 2005). Exposure therapy is based on the theoretical principle of extinction of conditioned fear responses (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014), whereby a conditioned stimulus (CS) is repeatedly presented in the absence of the aversive event until fear gradually declines. Although exposure-based CBT has been found to be effective in alleviating children's anxiety in approximately 60% of anxious youth, 40% are unlikely to benefit either in the short- or long-term, highlighting the need for continued basic research on extinction in youth (Ginsburg et al., 2014).

Most often, research on fear acquisition and extinction has employed a Pavlovian differential fear conditioning and extinction paradigm. During the differential conditioning phase, a neutral conditioned stimulus (CS+) is paired with an aversive unconditioned stimulus (US) so that the CS + acquires the capacity to elicit fear, whereas another conditioned stimulus (CS-) is never paired with the aversive US (Duits et al., 2015). Differential fear conditioning is typically observed by an increase in self-reported and physiological fear-related responses to the CS + compared to the CS-. During the extinction phase in which both the CS+ and CS- are repeatedly presented, the US is no longer presented with the CS + trials. Extinction is indexed by decreases in self-

* Corresponding author. School of Applied Psychology, Griffith University,Queensland, 4222. Australia.
** Corresponding author. School of Applied Psychology, Griffith University, Queensland, 4222. Australia.
E-mail addresses: k.ryan@griffith.edu.au (K.M. Ryan), a.waters@griffith.edu.au (A.M. Waters).

https://doi.org/10.1016/j.brat.2018.11.009
Received 26 October 2017; Received in revised form 27 October 2018; Accepted 19 November 2018
Available online 20 November 2018
0005-7967/

reported and physiological responses to the CS + to within comparable levels to the CS-. Many studies also examine relapse of fear via mechanisms including reinstatement, renewal and spontaneous recovery. Reinstatement occurs when extinguished fear returns after exposure to an aversive stimulus; renewal takes place when the CS is presented in a different context to the extinction context, and spontaneous recovery occurs following a period of time without encountering the CS or US (for a review, see Boschen, Neumann, & Waters, 2009). Studies in which relapse of fear are examined include an additional phase in which the CS+ and CS- are presented and these trials may occur shortly or much later (e.g., on the next day) following extinction.

This review was triggered by the observation of variation in results from studies of differential fear conditioning and extinction among children and adolescents. This variation may be due to differences associated with cognitive, biological and emotional development (Rapee, Schniering, & Hudson, 2009; Steinberg, 2011; Waters & Craske, 2016). However, findings may have also been influenced by variation in task design, methodology, and the dependent measures assessed. For example, researchers have utilised different types of CSs and USs, different numbers of trials during differential conditioning and extinction phases, and differing schedules of reinforcement on CS + trials during conditioning. Moreover, only a handful of studies in children and adolescents have included extinction retest phases or examined mechanisms underlying return of fear following extinction. There are also other challenges facing this area of research. For example, some fear-evoking stimuli are unacceptable to use with youth, which limits the options available to researchers. Furthermore, some studies involve stimulus parameters and task designs that are more complex and require active participant involvement during the experiments (e.g., providing trial-by-trial US expectancy and CS evaluation ratings) whereas others involve passive task conditions during differential fear conditioning and extinction and assess between phase ratings only. The extent to which variations in task design and methodology invoke differing levels of cognitive demand upon children relative to adolescents and contribute to mixed results remains unclear. Furthermore, determining the design and methodological parameters of differential conditioning and extinction paradigms that produce theory-consistent and replicable effects is important for providing a solid experimental framework to examine mechanisms in the laboratory that can then be translated into intervention research on exposure-based treatments (Waters, LeBeau, & Craske, 2017). A systematic review is an important first step towards identifying task parameters that may be most reliable in producing differential fear conditioning and extinction effects in experimental research with children and adolescents.

1.1. Literature review of design features

1.1.1. Design

Single cue conditioning refers to the presentation of a single CS paired with the US. Responding to the CS is compared with a control group that receives random or explicitly unpaired presentations of the CS and the US. The single cue design has been criticised for failing to control for orienting and other non-associative processes that may affect responding to the CS (Luck & Lipp, 2016). Also, finding appropriate controls is difficult as is the use of an explicitly unpaired stimulus sequence as it can result in inhibitory conditioning (when the CS signals the absence of the US) to the CS (Luck & Lipp, 2016). In contrast, differential fear conditioning uses two CSs, one that is paired with the US (CS+) and another that is not (CS-). A conditioned response (CR) is elicited by the CS+ and not by the CS-, thereby controlling for non-associative factors through the use of a within-participant design, and as such, it provides a statistically more powerful approach to examine conditioning than single cue designs. As differential cue conditioning is recommended over single cue conditioning (Lissek et al., 2005), only differential fear conditioning and extinction studies are reviewed in this paper.

1.1.2. Type of CS

Conditioned stimuli (CS) are used to signal the presence of the US in differential conditioning procedures. The type of CS has varied extensively across research studies; however, differential conditioning is most robust when the CS type is neutral before beginning acquisition training, that is, it will only elicit a differential conditioned response after repeated pairing with the US (Lipp, 2006; Lissek et al., 2005). A meta-analysis of adult studies found the most common types of CSs were visual and auditory stimuli, although a few studies have used odour CSs (Shechner, Hong, Britton, Pine, & Fox, 2014).

Auditory stimuli such as tones of varying quality and intensity were used frequently in adult studies to differentiate between the CS+ and CS- particularly in the 1950 to 1960's (Howe, 1957; Fayu, 1961; Clum, 1969; Ashcroft, Guimarães, Wang, & Deakin, 1991; Del-Ben et al., 2001). In youth studies, auditory CSs have rarely been used (Block, Sersen, & Wortis, 1970; Gao, Raine, Venables, Dawson & Mednick, 2010 a, b, c and Gao, Tuvblad, Schell, Baker, & Raine, 2015) primarily because differentiating between two sounds could be difficult for youth. Moreover, none of the studies with children tested contingency awareness or US expectancy, so it is unclear whether participants correctly identified the CS + -US association, and extinction was not reported.

Odour CSs have primarily been used in studies of patients with psychosomatic conditions to provoke hyperventilation as the CR (Van den Bergh, Stegen, & Van de Woestijne, 1997). There appears to be no studies that have used odour CSs with youth. Moreover, odour perception is thought to be related to disgust rather than fear and can evoke memories or emotions of prior learning that can affect individuals differently (Sehlmeyer et al., 2009). Also, ethically it may be difficult to use odour as the CS with youth, as it may induce illness and distress, as found in a study with adults in which common symptoms included fatigue, difficulty concentrating, pounding heart, shortness of breath, anxiety and headache (Van den Bergh et al., 1997).

Of visual stimuli, word CSs were used in an early study on adult patients with varying phobias (psychasthenic outpatients), and although the results found differential conditioning and extinction in all participant groups, there was no significant differences between groups (Halberstam, 1961). Halberstam (1961) suggested that words could induce different memories and meanings for individuals and therefore may not be considered neutral. Moreover, word CSs requires a level of reading ability, which may be problematic when studying differential fear conditioning and extinction in young children. There appears to be no studies using word CSs with youth.

Face CSs have been used in studies with youth. The features of faces have varied widely amongst studies with adults and youth, with variations observed in face gender, emotion and race; for example, neutral faces (Lau et al., 2008), angry faces (Rowles, Lipp, & Mallan, 2012), male (Tzschoppe et al., 2014) and female faces (Chauret et al., 2014) as well as those of different race, e.g., Asian (Mallan, Sax, & Lipp, 2009). Although face CSs are thought to be evolutionarily-relevant and suitable for eliciting fear processes akin to those associated with phobic anxiety (Ohman & Mineka, 2001), they may be more meaningful to participants compared to other neutral stimuli such as geometric shapes, and thus, susceptible to a classical conditioning phenomenon known as latent inhibition, which states that familiar stimuli result in weaker conditioning than do unfamiliar and neutral objects (Mineka & Oehlberg, 2008; Sehlmeyer et al., 2009). This is due to the amount of pre-exposure humans have to faces, although it should be noted that children may be equally exposed to geometric shapes on a regular basis due to educational play. Also, a study has found gender effects when using four male and four female face CSs, with female participants eliciting stronger amygdala activation than male participants (Williams et al., 2005). Furthermore, like odours and words, face CSs could trigger strong emotions and memories associated with prior events and thus the potential for face CSs to be more resistant to extinction than more neutral CS types such as shapes (Lissek et al., 2005).

Indeed, shape CSs have been used extensively in studies with adults and youth and have included different coloured squares (Pattwell et al., 2012) or different coloured geometric shapes such as triangles and circles (Waters, Henry, & Neumann, 2009) and also more complex shapes such as cartoon-like faces (Liberman, Lipp, Spence, & March 2006) and three dimensional animal like combinations of shape and colour similar to real world animals (Barry, Vervliet, & Hermans, 2016).

1.1.3. Type of US

The unconditioned stimulus (US) is used to induce fear of the conditioned stimulus (CS+) and the challenge in youth research has been to employ sufficiently fear-eliciting stimuli that are safe and acceptable for children. Research with adult participants has traditionally used electric shocks as the aversive US (Lissek et al., 2005). However, as shock is unethical to use with youth because it may invoke distress and discomfort, and it requires insight from the participant during the shock work up procedure to determine a level that is uncomfortable but not painful, novel but less aversive stimuli have been employed in youth studies. Researchers have primarily used auditory stimuli such as keys rattling in a tin can (Gao et al., 2015), pure tones (Craske et al., 2008), white noise (Fairchild, Van Goozen, Stollery, & Goodyer, 2008), metal fork scraping on slate (Neumann, Waters, & Westbury, 2008), and human screams (Lau et al., 2008).

An air puff US has been used in adult and adolescents studies and involves an unpleasant air blast directed at the larynx. Grillon et al. (1999) used air blasts as aversive startle eliciting stimuli during fear-potentiated startle experiments in adolescents and adults, though they did not test extinction. Pine et al. (2001), found differential conditioning effects as indexed by amygdala activation during fMRI experiments with adults when using an air puff US and different coloured lights as the CS to measure changes in amygdala activity. In a differential fear conditioning and extinction study of trauma-exposed adults with and without PTSD, successful differential fear conditioning but not extinction was observed using an air blast US to the larynx and different coloured shapes CSs (Norrholm et al., 2011). An adolescent study found that healthy participants rated an aversive air blast to the larynx as mildly fearful, and therefore concluded that it might be suitable for studying differential fear conditioning and extinction in anxious patients (Monk et al., 2003). Reeb-Sutherland et al. (2009) used an air puff US and geometric shapes CSs during a differential conditioning study with behaviourally inhibited adolescents with and without anxiety measured by FPS which was elicited by a white noise startle eliciting stimulus. They found all groups showed significant differential fear conditioning. Jovanovic et al. (2014) demonstrated successful extinction in children using an airblast US and white noise as the blink eliciting stimulus.

The use of odour as a US has also been examined and requires tubes to be inserted in the nostrils or a mask that covers the mouth and nose, to deliver the smell effectively. The type of odour used has varied among studies. Adult studies have used odour US with successful yet weak differential conditioning results on healthy controls (Hermann, Ziegler, Birbaumer, & Flor, 2002). Van den Bergh et al. (1997) used odour as a CS, and CO_2 as the US; the CO_2 was used to trigger hyperventilation in adult psychosomatic patients. Another study used odour US and observed differences in amygdala changes in phobic patients relative to controls during conditioning but not extinction with adults (Schneider et al., 1999). Odour USs have not been used with youth, possibly because the method of delivering the odour may be distressing for children.

The human scream US has been used in numerous adult and adolescent studies. Developed by Lau et al. (2008), this US typically involves a female scream ranging from 80 dB to 95 dB and it is paired with a fearful face for durations of 100 ms–300 ms (Britton et al., 2013; Glenn et al., 2012; Lau et al., 2011). Original testing with youth using self-report responses demonstrated that the CS + face paired with the scream US was significantly more fear provoking than the CS- face (Lau et al., 2008). However, more than half the participants dropped out of the experiment before the day two extinction phase which authors attributed to the high fear level (as participants returned to partake in other studies). The researchers concluded that the US scream was more aversive than tones and air puffs, but less aversive than electric shocks, suggesting that it was an effective mechanism to use in differential fear conditioning and extinction studies with youth (Lau et al., 2008; Lissek et al., 2005). Given the potential for high dropout rates in studies with adolescents (Britton et al., 2013) and the fact that the scream US may be too aversive with children, Shechner et al. (2015) developed a novel bell US paradigm to address the limitations of the aversive scream US. The results revealed much lower dropout rates based on a smaller sample size than that of prior studies (e.g., Britton et al., 2013).

Several studies have used pure tones at 1000 Hz delivered through headphones, which create an unpleasant direct experience. However, discrepancies exist in the intensity and duration of tones e.g., 1000 Hz at 107 dB (Craske et al., 2008; Waters et al., 2009), 1000 Hz at 105 dB (Liberman et al., 2006). White noise is a high frequency sound typically set at around 105 dB with a duration of 100 ms–600 ms (Lissek et al., 2005; Peri, Ben-Shkhar, Orr & Shalev, 2000; Pliszka, Hatch, Borcherding, & Rogeness, 1993). One adult study by Peri, Ben-Shakhar, Orr, and Shalev (2000) consisted of bursts of 105 dB 500 ms white noise USs, finding that those with PTSD had larger heart rates than healthy controls during differential conditioning and reduced extinction of conditioned responses. A child study (6–12 years) conducted in the early 90's used shapes as the CS and white noise (110 dB 200 ms) as the US, differential conditioning occurred for all groups (ADHD, anxious and healthy) (Pliszka et al., 1993). The researchers concluded that a ceiling effect may have occurred because the US was of such high intensity; but later refuted this conclusion, suggesting that most children appeared bored with the experiment, and concluded that they observed habituation to the US (Pliszka et al., 1993). However, white noise is generally not recommended for children, the elderly, those with hearing problems, and especially infants and young children who can be sensitive to the high intensity acoustic stimuli, particularly when used in repeated exposure (Lissek et al., 2005).

A novel ecological sound was created by Neumann and Waters (2006), which consisted of a 3-s recording of a metal fork scraping over slate. This stimulus was rated as more unpleasant than an electric shock by adult participants and has been shown to produce reliable differential fear conditioning and extinction results in adults (Neumann & Waters, 2006), adolescents (Neumann, Waters & Westbury, 2008) and children (Neumann, Waters, Westbury, & Henry, 2008).

1.1.4. Number of conditioning trials

There has been wide variation in the number of trials used during differential conditioning and extinction phases. Discrepancies are often due to the purpose of the research, i.e., to establish reliable differential conditioning to examine the effects of extinction or fear generalisation to other stimuli. Furthermore, the number of trials may simply be based on researcher training and experience. In adult studies, the number of trials has ranged from 3 to 60 for conditioning and 2 to 25 for extinction (Hermann et al., 2002; Lissek et al., 2005; Van den Bergh et al., 1997). The number of trials has been considerably smaller in studies with youth and has typically ranged between 6 and 24 for conditioning and 3 to 24 for extinction (Pattwell et al., 2012), with one study including 60 trials during conditioning (Lau et al., 2011). The choice of fewer trials with youth may circumvent concentration problems and fatigue and limit exposure to the unpleasant US in younger participants.

1.1.5. CS-US reinforcement schedule

The frequency with which the US is paired with the CS has also varied widely across adult and youth studies, and has largely been based on the desired effects during extinction. The decision to use partial reinforcement of either 50% or 75%, instead of 100% reinforcement during differential conditioning, is to prevent habituation

of responses to the US and so leading to more resistance during extinction (Lau et al., 2008). Research that investigated neural reactivity to 50% and 100% CS-US reinforcement showed the largest amygdala responses during the 100% trials, because the amygdala is the principal site of CS-US convergence and conditioned response production during differential fear conditioning (Dunsmoor, Bandettini, & Knight, 2007). Accordingly, *not* using 100% CS-US reinforcement affects participants' CS-US contingency awareness, which in turn, slows the extinction of fear responses (Chauret et al., 2014). Thus, 100% reinforcement produces greater CS-US contingency awareness during differential conditioning but also produces more rapid extinction than partial CS-US reinforcement schedules during differential conditioning.

The purpose of this review was to systematically evaluate the experimental designs, methodological features, and measurement parameters of differential fear conditioning and extinction studies conducted with youth to (a) identify the most reliable experimental parameters and dependent measures for extinction studies, and (b) formulate recommendations for future research. In this review, we focus only on the results of healthy participants to reduce any confounding factors, however search terms included "anxiety" to assist in finding differential fear conditioning and extinction studies, as the examination of these processes in anxious individuals has often been the purpose of these experiments. In this review, we describe features of differential fear conditioning and extinction studies, including the type of CS, the type of US, the number of CS conditioning and extinction trials and the CS-US reinforcement schedules used in prior studies. In addition, we examine the measurement parameters of skin conductance responses, fear potentiated startle, and subjective report measures (e.g., US expectancy, CS evaluations and subjective anxiety ratings). Contingency awareness, another subjective report measure of the extent to which the participant has established the association between the CS + and the US, is discussed separately.

2. Method

2.1. Inclusion & exclusion criteria

Fig. 1 displays the search strategy. Articles were included if they met the following initial search parameters: (a) published between 1970 to July 2017; (b) written in English; (c) tested fear conditioning; (d) contained healthy participants either as the only participants or control group, and (e) included children or adolescent humans between 2 and 17 years of age.

2.2. Search strategy

Searches were conducted using Medline, Embase and PsycInfo databases. Search terms included (child* OR adolesce* OR youth) AND (Fear conditioning) AND (differential cond*) AND (Acquisition) AND (Extinction) AND (anxiety or anxious). Plus, MeSH terms – Anxiety; Extinction, Psychological; Conditioning, Psychology; conditioned stimulus; conditioned fear. These were searched with keyword parameters. All papers up until and including 21st July 2017 were included in the search.

2.3. Screening

Parameters of the results were refined to exclude non-articles (e.g. Posters, letters, lists of abstracts), publications not in English and duplication across databases. For the second exclusion, the keywords (rats and mice) were entered into EndNote X8 to eliminate those that did not fit the criteria for human participants. In the third exclusion, only articles that contained the keywords child, children, adolescent or adolescence or youth were retained. For the remaining articles, titles and abstracts were screened for inclusion or exclusion criteria suitability, in particular that they were differential fear conditioning studies. For the final review, the full text was studied to determine if they contained normal or healthy control groups in the experiments.

2.4. Assessment method for establishing the quality of differential fear conditioning and extinction experiments

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, which consists of a 27-item checklist, was used to ensure transparent reporting of studies included in this review (Moher, Liberati, Tetzlaff, Altman, & Group, 2009). In addition, the assessment guidelines used for evaluating the quality of measures in the Effective Public Health Practice Project (EPHPP) (Thomas, Ciliska, Dobbins, & Micucci, 2004) were adapted to develop a Differential Fear Conditioning and Extinction Quality Assessment Instrument (DFCE-QAI). The research team devised a list of component criteria to be included in the assessment process to identify task parameters and dependent measures used in differential fear conditioning and extinction experiments. The components included (a) design, (b) type of CS, (c) type of US, (d) number of trials for conditioning and extinction and (e) percentage of reinforced trials (see Table 1). The first author developed a definition for each classification level of each component based on the terms and task parameters reported in the literature, with levels ranging from *strong*, *moderate* and *weak* based on the EPHPP guidelines (Thomas et al., 2004). The classification levels were applied in relation to the suitability of each component for use in differential fear conditioning and extinction experiments with children and youth rather than the relative strength of differential conditioning and extinction reported in each study. The definitions and classification levels were revised and ratified between the first and fourth authors until consensus was achieved. The definitions of each classification level for each component are defined below (see Table 1).

1. Design – For a *Strong* design classification, only differential or discrimination conditioning designs were included, based on the review by Lissek et al., (2005). A study needed to comprise of at least two CSs, one that was paired with an aversive US (CS+) and another that was not (CS-). To be classified as *Moderate* the design used had to be single cue conditioning, which used only one CS. To be classified as *Weak* the type of design was either not specified or a mix of other designs was used that were not focused on differential fear conditioning.
2. Type of CS – a CS was classified as *Strong* if it was (a) perceptually clear and distinct from the other CS, (b) free of evaluative properties, so as not to evoke emotions or memories that might engage greater cognitive processing and (c) was dissimilar to the US. Different modalities are recommended to reduce the potential confusion in the child and help maintain their attention to the task (Neumann, Waters, & Westbury, 2008). For a *Moderate* classification, the CS types were unclear and distinct from the US; the CS may also cause the participant to think and process the CS type, evoking a cognitive load. Similarities between the two CSs classified this component to be *Weak* as similarities may defer a young participant's attention from the task.
3. Type of US – a US was classified as *Strong* if it was potent or aversive based on combined rater evaluation in relation to the suitability for use in differential fear conditioning and extinction studies with children and youth. To be aversive, stimuli are defined as resulting in mild pain or discomfort that is often associated with biologically harmful or damaging substances or events (Colman, 2009). To be classified as *Strong* the US needed to (a) be directly unpleasant in duration and/or decibel level, (b) have pre-existing emotional valence, and (c) be distinctly different to the CSs. A *Strong* rating example was the 'scream and fearful face' US developed by Lau et al. (2008) as this US conveys pre-existing emotional valence that loud tones lack. For the *Moderate* rating, the US was not evaluated as sufficiently aversive and showed some similarities to the CSs. To be

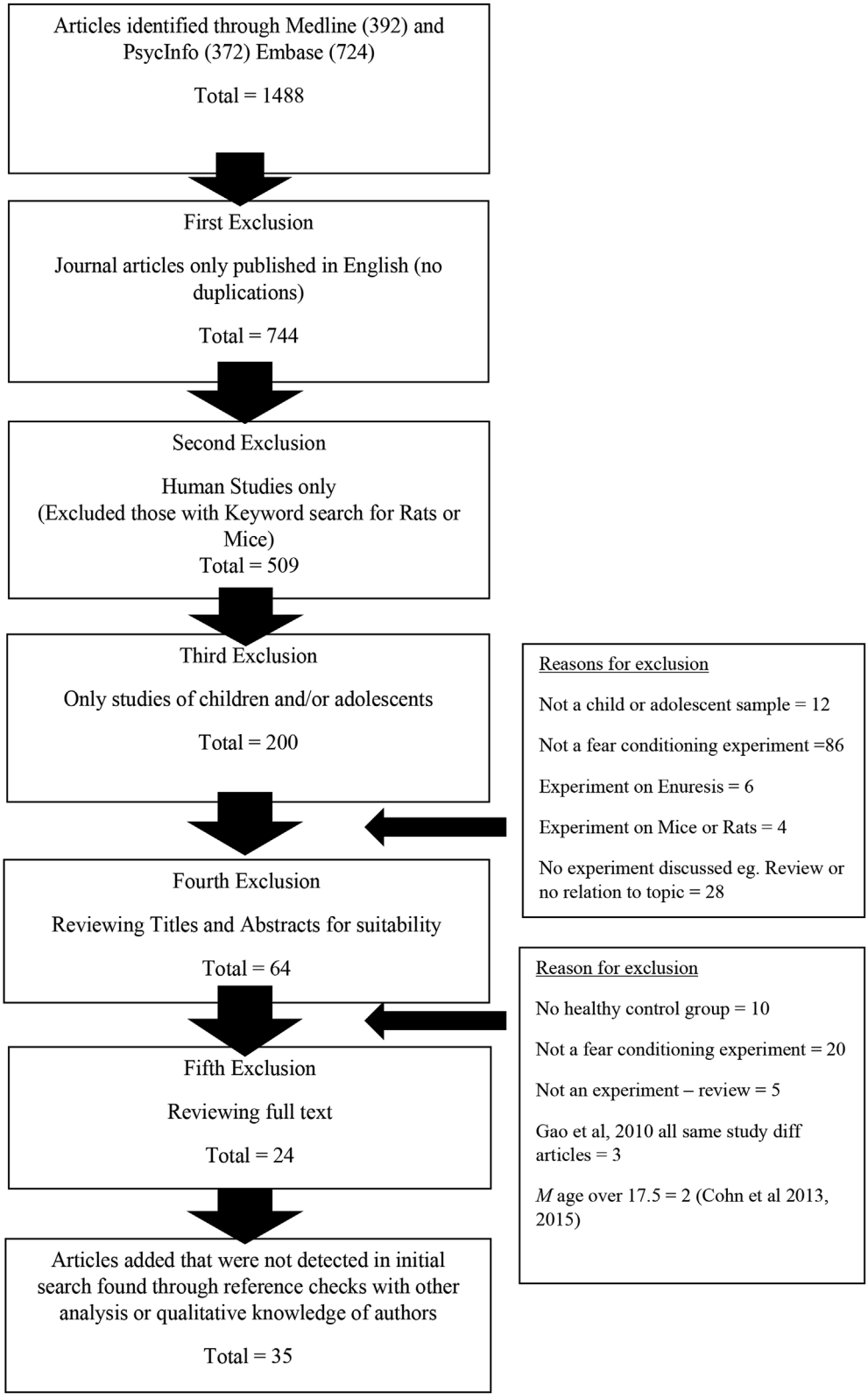


Fig. 1. Systematic review exclusion chart.

classified as *Weak,* the US type was not aversive or was similar to the CSs.

4. Number of Trials in conditioning and extinction – In differential conditioning, to be rated as *Strong*, the number of trials of the CS was classified as 8 or more. This was supported by Neumann et al. (2008) who recommended when assessing SCR in experiments with children that a minimum of 8 and up to 12 trials per CS should be used, as differential conditioned responses were evident only by trial 8. Liberman et al. (2006) found that only 6 conditioning trials of each CS was insufficient to support differential conditioning in children. *Moderate* ratings were classified as 4–7 trials of the CS as researchers found consensus on this number. Less than 4 trials were

Table 1
Differential fear conditioning and extinction quality assessment instrument (DFCE-QAI) definitions.

Components	Strong	Moderate	Weak
Design	Differential conditioning design	Single cue conditioning, only use the CS+	Unspecified or mixed with other experiments
Type of CS	1 Perceptually clear and distinct. 2 Neutral of evaluative properties, so as not to evoke emotions or memories that might engage greater cognitive processing 3 Dissimilar to the US	Not clear and distinct or would evoke a participant to think or evaluate the item	Close similarities between the CS+ and CS-
Type of US	The US is potent or aversive and not similar to CS's 1 Directly unpleasant – Duration; dB(A) 2 Has pre-existing emotional valence (evokes positive or negative emotions) 3 Distinctly different to the CSs	Neither very aversive or very similar to CS's	Close similarities to the CS's and not aversive
No of trials – how many of each CS is shown in ACQ	8 or more	4–7	3 or less
No of trials – how many of each CS is shown in EXT	5 or more	4	3 or less
% of reinforced trials	100%	50%–99%	Below 50%

Note. CS – Conditioned stimulus; US – Unconditioned stimulus; ACQ – Acquisition; EXT – Extinction.

classified as *Weak*. For extinction, the number of trials for a *Strong* classification was adopted as 5 or more trials per CS. This was supported by past research that discussed their limitations for achieving stronger results when fewer trials were used in extinction (Waters et al., 2009). For *Moderate* ratings, 4 trials was deemed reasonable based on researcher consensus. For *Weak* ratings, 3 or fewer trials of the CS were considered unsatisfactory for successful differential fear conditioning and extinction designs.

5. Percentage of reinforced trials – As the focus in this review was on designs from which extinction effects were observed, the *Strong* rating for the percentage of reinforced trials between CS-US was 100% as total reinforcement was hypothesised as the most effective for differential fear conditioning and extinction results. The *Moderate* rating was decided by researchers as 50%–99% to encapsulate those partial reinforcement studies with the purpose of being more resistant to extinction. The *Weak* rating was given to any studies in which a 50% or less reinforcement schedule was used, as differential fear conditioning would be even more resistant to extinction.

Two independent raters evaluated the studies by rating each experimental component in Table 1 using these classification levels described above. Afterwards, the raters met with the first author during a consensus meeting, they presented their rating of each experimental component and ratings were classified as meeting consensus if the same classification rating was given by both raters. Any differences were discussed until consensus was reached and the raters' were questioned about their interpretation of the definition or classification level to determine if any amendments were required to enhance clarity. The interrater reliability level of agreement was 83%.

3. Results

3.1. Article selection

The process of inclusion and exclusion of articles is summarised in Fig. 1. Initially 1488 articles were identified and after review of keywords, titles and abstracts, 64 articles were fully reviewed. The reasons for article exclusion are listed in Fig. 1. After full article review, another 40 articles were excluded (reasons are described in Fig. 1). One longitudinal study (Gao et al., 2010b, 2015, 2010c, 2010a) was documented in four articles, and therefore remained in the review as the 2015 article. After cross checking reference lists and previous reviews, as well as author qualitative knowledge, 9 articles were added to the review. Two papers report two experiments (Lau et al., 2011; Shechner et al., 2015), so they were added as four separate studies in the final total. This resulted in a total of 35 included studies (all included references are highlighted with * in reference list).

Experimental parameters and dependent variables of the 35 studies identified were decomposed and are presented in Table 2 including the number of participants, the mean age and age range. The type of CS and US are detailed in Table 2, with descriptions of the picture or sounds used, along with the intensity of tones and length of exposure to the US. The number of trials for each CS at conditioning and extinction is listed plus the percentage of those trials that were reinforced by the US. The duration of the CS and intertrial intervals (ITIs) were also listed for comparison purposes.

The differential fear conditioning and extinction quality assessment instrument criteria (as per Table 1) were then applied to each paper, the resulting table represents consensus ratings from independent raters (see Table 3). All studies were classified as having a strong design, i.e., all studies identified in this review were differential conditioning designs and there were no single cue designs. The rating of strength of suitably of type of CS and US was a subjective evaluation despite detailed criterion. Subjective decisions included: faces (CS) as a moderate rating as faces may evoke emotions or memories causing cognitive load that may influence responding, and the same face appears with a different expression forming part of the US. The geometric shapes are unlikely to invoke prior memories and learning and were rated as strong. The scream (US) was rated as strong, due to its aversive and fear-provoking properties. The tones (US), such as the pure tone, white noise and metal scrapping on slate; were rated as moderate in terms of aversiveness and fear-provoking properties.

3.2. Dependent measures

Table 4 details the dependent measures used in each study. The main measures used were skin conductance response (SCR), fear potentiated startle (FPS) and subjective report (SR) measures (See supplementary paper for review of dependent measures). The other measurements identified included heart rate, neural measures, eye movements and reaction-time to index avoidance. The units of measurement in each study are outlined in Table 4.

3.2.1. Skin conductance response

Of the 35 studies, 23 assessed SCR. SCRs serve as an objective, nonverbal, involuntary indicator that is sensitive to the anticipation of a possible significant outcome (Dawson, Schell, & Courtney, 2011). The most popular unit of measurement (11 studies) was the whole interval response, which is defined as using either magnitude or amplitude SCR

Table 2
Design paradigms.

Authors	No. of Participants and *M*(Age) [Age Range]				Type of CS	Type of US	No of trials (Each CS)		% of reinforced trials	Duration of CS	Duration of ITI
	ACQ		EXT				ACQ	EXT			
	Controls	ANX	Controls	ANX							
Block et al. (1970)	77 [2–11]	Downs 24 [4–11] Brain trauma 29 [6–11]	NA	NA	CS+ 400 Hz tone CS- 1000 Hz tone (55 dB)	Auto horn (1s) 95 dB	10	NA	100%	5 s	20 s
Pliszka et al. (1993)	22 (M = 10.2) [6–12]	11 (M = 9.9) 23 ADHD (M = 9.5) [6–12]	22 (M = 10.2) [6–12]	11 (M = 9.9) 23 ADHD (M = 9.5) [6–12]	CS + Red square with 1000 Hz tone CS- Blue square with 500 Hz tone	White noise blast (0.2s) 110 dB	8	4	100%	4 s	NR
Liberman et al. (2006)	30 (M = 9.7) [7–13.5]	53 (M = 10.1) [7–14]	30 (M = 9.7) [7–13.5]	53 (M = 10.1) [7–14]	Custom designed cartoons	1000 Hz Pure tone at (500 ms) 105 dB	6	8	100%	5 s	15, 20, 25 s
Craske et al. (2008)	11 (M = 9.4) [7–12.75]	23 (M = 9.4) 15 At risk [7–12.75]	11 (M = 9.4) [7–12.75]	23 (M = 9.4) 15 At risk [7–12.75]	Geometric Shapes Pink trapezoid and pastel cream triangle	1000 Hz pure tone at (1s) 107 dB	8	4	100%	8 s	20, 25, 30 s
Fairchild et al. (2008)	54 (M = 15.8) [14–18]	71 CD (M = 15.8) [14–18]	54 (M = 15.8) [14–18]	71 CD (M = 15.8) [14–18]	Coloured slides	White noise tone (1s) 99 dB	18	6	55%	3 s	10 s
Lau et al. (2008)	38 (M = 13.9) [11.7–16.3]	16 (M = 13.6) [10.4–15.3]	38 (M = 13.9) [11.7–16.3]	16 (M = 13.6) [10.4–15.3]	Faces	Fearful face (3s) + scream (3s) 95 dB	16	15	75%	8 s	NR
Neumann et al. (2008a)	15 (M = 16.0) [13–17]		15 (M = 16.0) [13–17]		Geometric shapes Diamond and triangle	Recording of 3-pronged fork scraped on slate (3s) 83 dB	8	8	100%	8 s	20, 25, 30 s
Neumann et al. (2008b)	16 (M = 9.6) [8–11]		16 (M = 9.6) [8–11]		Geometric Shapes Small white square and large black square	Recording of 3-pronged fork scraped on slate (3s) 83 dB	12	12	100%	8 s	13–16 s
Reeb-Sutherland et al. (2009)	51 (M = 15.5) [13.1–17.3]	25 (M = 15.5) [13.1–17.3]	NA	NA	Geometric shapes Grey squares and white squares	Air blast 700 kPa	8	NA	50%	4 or 7 s	10, 20 s
Waters et al. (2009)	18 (M = 10.3) [8–12]	17 (M = 10.2) [8–12]	18 (M = 10.3) [8–12]	17 (M = 10.2) [8–12]	Geometric shapes Pink trapezoid and cream triangle	1000 Hz Pure tone (1s) 107 dB	8	4	100%	8 s	20, 25, 30 s
Fairchild et al. (2010)	55 (M = 15.3) [14–18]	25 (M = 15.6) [14–18]	55 (M = 15.3) [14–18]	25 (M = 15.6) [14–18]	Coloured slides	White noise tone (1s) 97 dB	9	6	56%	3 s	10 s
Gao et al. (2015)	3yrs n = 200 4yrs n = 195 5yrs n = 190 6yrs n = 184 8yrs n = 172		NA	NA	2 tones CS+ 1000 Hz CS- 500 Hz (12.5s) 60 dB	White noise -Keys in a tin can (4.5s) 90 dB	4, 8 and 12	NA	66%	10 s	38 s
Haddad et al. (2011)	42 (M = 13.5) [12–15]		42 (M = 13.5) [12–15]		Three Faces	Face with expression and comments, negative, positive and neutral	9	8	75%	1 s	1 s
Lau et al. (2011) EXP 1 EXP 2	21 (M = 13.1) [10.1–16.0] 21 (M = 27.1) 15 (M = 13.3) [11.0–15.7] 20 (M = 28.9)		NA NA		Faces	Fearful face (3s) + scream (1s) 90 dB Fearful face + scream (1.1s) 90 dB	10 60	NA NA	80% 50%	8 s 7.1 s	8–10 s 2–4 s
Glenn et al. (2012)	40 (M = 10.8) [8–13]		NA		Faces	Fearful face (3s) + scream (1s) 80 dB	8	NA	75%	6 s	9–11 s
Patwell et al. (2012)	30 (M = 8.8) [5–11] 28 (M = 13.9)		30 (M = 8.8) [5–11] 28 (M = 13.9)		Geometric shapes - Coloured squares	1000 Hz White noise (1s) 85–107 dB	12 (Day 1)	24 (Day 2)	50%	3 s	13 s

(*continued on next page*)

Table 2 (*continued*)

Authors	No. of Participants and *M*(Age) [Age Range]				Type of CS	Type of US	No of trials (Each CS)		% of reinforced trials	Duration of CS	Duration of ITI
	ACQ		EXT				ACQ	EXT			
	Controls	ANX	Controls	ANX							
	[12–17] 25 (*M* = 22.8) [18–28]		[12–17] 25 (*M* = 22.8) [18–28]								
Britton et al. (2013)	42 (*M* = 14.2) [8–17] 31 (*M* = 28.8)	23 (*M* = 13.3) [8–19] 18 (*M* = 32)	42 (*M* = 14.2) [8–17] 31 (*M* = 28.8)	23 (*M* = 13.3) [8–19] 18 (*M* = 32)	Faces	Fearful face (1s) scream (1s) 95 dB	10	8	80%	7–8 s	11–15 s
Chauret et al. (2014)	117 (*M* = 13.9) [10–17]		117 (*M* = 13.9) [10–17]		Male and female faces	Fearful face and scream (1.1s) 90 dB	28	14	50%	3 s	3–12 s
Jovanovic et al. (2014)	27 (*M* = 10.3) [8–13]	33 (*M* = 10.3) [8–13]	27 (*M* = 10.3) [8–13]	33 (*M* = 10.3) [8–13]	Geometric shapes Blue square and purple triangle	Air blast to larynx (80psi)	9	12	100%	6 s	9–22 s
Waters et al. (2014)	31 (*M* = 9.6) [8–11]	14 Anx (*M* = 9.8) 12 Dep (*M* = 9.9) [8–11]	31 (*M* = 9.6) [8–11]	14 Anx (*M* = 9.8) 12 Dep (*M* = 9.9) [8–11]	Geometric shapes Pink trapezoid cream triangle	1000 Hz pure tone (1s) 100 dB	8	4	100%	8 s	20–30 s
Tzschoppe et al. (2014)	47 (*M* = 14.0)		47 (*M* = 14.0)		Male faces	Female scream (1s) 80 dB	15	10	50%	6 s	8–10 s
Den et al. (2015)	70 (*M* = 14.8) [12–17] 48 (*M* = 47.8)		70 (*M* = 14.8) [12–17] 48 (*M* = 47.8)		Faces	Fearful face (3s) + scream (3s) 75 dB	7	5	71%	8 s	10 s
Haddad et al. (2015)	11 (*M* = 15.6) [14–17.6]	15 (*M* = 15.2) [12.8–17.6]	NA	NA	Faces	Fearful face and scream (1s) 95 dB	30	NA	50%	3 s	2–4 s
Kadosh et al. (2015)	24 (*M* = 14.0) [12.9–15.3]	19 (*M* = 14.6) [13.2–16]	NA	NA	Faces In 3 virtual room contexts	Face and scream (750 ms) 95 dB	24	NA	100%	2 s	1–5 s
Shechner et al. (2015)	17 (*M* = 13) [10–16] 30 (*M* = 29.7)	15 (*M* = 11.5) [8.9–14] 22 (*M* = 29.2)	17 (*M* = 13) [10–16] 30 (*M* = 29.7)	15 (*M* = 11.5) [8.9–14] 22 (*M* = 29.2)	Faces Pictures of blue and yellow bells	Fearful face (1s) scream (1s) 95 dB Red bell with aversive bell sound (1s) 95 dB	10	8	80%	7–8 s	8–21 s
McGuire et al. (2016)	22 (*M* = 12.6) [8–17]	OCD 19 (*M* = 13.3) [8–17]	22 (*M* = 12.6) [8–17]	OCD 19 (*M* = 13.3) [8–17]	Faces	Fearful face and scream (3s) 95 dB	10	8	80%	8 s	5 s
Waters and Pine (2016)	32 (*M* = 9.9) [7–13]	44 (*M* = 10.0) [7–13]	32 (*M* = 9.9) [7–13]	44 (*M* = 10.0) [7–13]	Geometric Shapes Pink trapezoid and cream triangle	1000 Hz pure tone (1s) 100 dB	8	4	100%	8 s	20–30 s
Schiele et al. (2016)	267 (*M* = 9.0) [8–10] 285 (*M* = 25.6)		NA		Faces	Fearful face and scream (1.5s) 95 dB	12	NA	82%	6 s	9–12 s
Michalska et al. (2016)	59 (*M* = 8.0) [5–10]		59 (*M* = 8.0) [5–10]		Picture of blue and yellow bells	Red bell with aversive bell sound (1s) 95 dB	10	8	80%	7–8 s	8–12 s
McLaughlin et al. (2016)	55 (*M* = 13.5) [6–18]	Maltreated 35 (*M* = 13.6) [6–18]	55 (*M* = 13.5) [6–18]	Maltreated 35 (*M* = 13.6) [6–18]	Picture of blue and yellow bells	Red bell with aversive bell sound (1s) 95 dB	10	8	80%	7–8 s	8–12 s
Newall, Watson, Grant, and Richardson (2017)	73 (*M* = 9.4) [7–12]		22 (*M* = 9.32) [7–12]		Animal pictures	Fearful human face	10	10	100%	1 s	2–4 s
Jackson et al. (2016)	77 (*M* = 13.2) [8–14]		NA		Female faces	Fearful face and scream	8	NA	75%		9–11 s
Waters et al. (2017)	46 (*M* = 8.8) [8–10] 35 (*M* = 16.1) [15–17] 42 (*M* = 32.3)		46 (*M* = 8.8) [8–10] 35 (*M* = 16.1) [15–17] 42 (*M* = 32.3)		Geometric shapes Pink trapezoid yellow triangle	Recording of 3-pronged fork scraped on slate (3s) 85 dB	12	12	100%	10 s	

(*continued on next page*)

Table 2 (*continued*)

Authors	No. of Participants and *M*(Age) [Age Range]				Type of CS	Type of US	No of trials (Each CS)		% of reinforced trials	Duration of CS	Duration of ITI
	ACQ		EXT				ACQ	EXT			
	Controls	ANX	Controls	ANX							
											15–25 s

Note. CS – Conditioned stimulus; US – Unconditioned stimulus; ITI – Intertrial Intervals; ACQ – Acquisition; EXT – Extinction; ANX – Anxious patients; Downs – Down syndrome patient; ADHD – Attention Deficit Hyperactivity Disorder; CD – Conduct disorder; DBD – Disruptive Behaviour Disorder.

across the entire duration in which the CS was presented. First interval responding is defined as responses starting within 1–4s of CS onset and reflects the initial signal value of the CS that is enhanced when the CS+ is paired with the US. Second interval responding is defined as responses starting within 4–7s of CS onset, and reflects the anticipation of the US. The third interval responses begin 7–11s following CS onset and provide a means to examine responses to the US on CS + trials and the effects of no US on CS- trials (Luck & Lipp, 2015; Prokasy & Kumpfer, 1973). Eleven studies reported variants of interval responding: 6 studies used 1st interval, 2nd interval or 3rd interval responses or a combination of two of these intervals; 5 studies reported all three intervals. One study reported assessing SCR but did not report how it was measured or any results (Lau et al., 2011).

3.2.2. Fear potentiated startle

Nine studies assessed FPS, which is most commonly measured through recordings of the contraction of the orbicularis oculi muscle beneath the eye, which is involved in the contraction of the eyelid in response to an intense startle-eliciting stimulus (Grillon, 2002). FPS requires the participant to be fitted with electromyographic (EMG) electrodes usually placed under the eye to measure their eye blink startle reflex, with the signal conducted to a recording device and requires a degree of participant preparation (Blumenthal et al., 2005). The intense startle-eliciting stimulus is most commonly an acoustic white noise burst of 100–110 dB(A) with an instantaneous rise time which can be presented either with loud speakers or headphones, or an air puff directed to the skin generally in the upper part of the face or larynx (Blumenthal et al., 2005). Results from the review indicated that the blink eliciting stimuli included white noise (6 studies) and air blast (3 studies). White noise duration varied with 5 of the 6 studies using 50 ms and one using 40 ms durations. The decibel level varied between 95 dB and 110 dB. Only Liberman et al. (2006) and Reeb-Sutherland et al. (2009) used the same decibel level, but varied in their measurement of the response; Liberman et al. (2006) measured from foot-point to the peak of the response whereas Reeb-Sutherland et al. (2009) measured the magnitude of response from a pre-startle stimulus baseline average to the peak. The three studies that utilised air blast as the blink eliciting stimuli measured from a pre-stimulus baseline to peak amplitude. The air blast was the same in all three studies: 40 ms in duration and 4–10 psi for pressure.

3.2.3. Subjective report measures

Twenty-five of the 35 studies used subjective report ratings, with most using more than one type of evaluation.

Six studies recorded US expectancy ratings, with variation occurring in whether expectancies were assessed during each CS trial (i.e., trial-by-trial) (4 studies), after the conditioning phase (1 study) or after conditioning and extinction (1 study). In this latter study, participants were asked to respond to a question using the computer keyboard for the US expectancy ratings after each phase. Two of the trial-by-trial studies used a computer keyboard for ratings and two used a dial and pointer device. The scales differed between studies in terms of number of points on the response scale from −100 to +100 (2 studies); 1–100%; −1 to +1; 0–8 and Yes or No.

The CS evaluations varied in terms of the dimension of the CS being assessed including valence (unpleasant - pleasantness) (9 studies), fear or scariness (14 studies), arousal (8 studies), happiness, perceived control, and interest (1 each respectively). Seven studies used the self-assessment mannequin (SAM) to obtain CS valence ratings, a user-friendly Likert scale with cartoon-like figures, the scales available are for valence, dominance and arousal (Lang, 1980). In these seven studies, variations occurred in the number of points on the Likert scale, two studies used 5-point scales with the remaining five using 9-point scales. The majority of CS evaluations (21 studies) were assessed before conditioning, after conditioning and after extinction. Three studies (2 x fear; 1 x valence) used trial-by-trial ratings after each CS presentation.

Table 3
Review of differential fear conditioning and extinction quality assessment instrument (DFCE-QAI).

Authors	Design	Type of CS	Type of US	No. of trials in ACQ	No. of trials in EXT	% of reinforced trials
Block et al. (1970)	Strong	Weak	Moderate	Strong	Strong	Strong
Pliszka et al. (1993)	Strong	Moderate	Moderate	Moderate	Moderate	Strong
Liberman et al. (2006)	Strong	Moderate	Moderate	Moderate	Strong	Strong
Craske et al. (2008)	Strong	Strong	Moderate	Moderate	Moderate	Strong
Fairchild et al. (2008)	Strong	Strong	Moderate	Strong	Moderate	Moderate
Lau et al. (2008)	Strong	Moderate	Strong	Strong	Weak	Strong
Neumann et al. (2008a)	Strong	Strong	Moderate	Moderate	Strong	Strong
Neumann et al. (2008b)	Strong	Strong	Moderate	Strong	Strong	Strong
Reeb-Sutherland et al. (2009)	Strong	Strong	Moderate	Moderate	NA	Moderate
Waters et al. (2009)	Strong	Strong	Moderate	Moderate	Moderate	Strong
Fairchild et al. (2010)	Strong	Strong	Moderate	Strong	Moderate	Moderate
Gao et al. (2015)	Strong	Weak	Moderate	Strong	NA	Moderate
Haddad et al. (2011)	Strong	Strong	Weak	Strong	Strong	Moderate
Lau et al. (2011)	Strong	Moderate	Strong	Strong	NA	Strong
Glenn et al. (2012)	Strong	Moderate	Strong	Moderate	NA	Strong
Patwell et al. (2012)	Strong	Strong	Moderate	Strong	Strong	Moderate
Britton et al. (2013)	Strong	Moderate	Strong	Strong	Strong	Strong
Chauret et al. (2014)	Strong	Moderate	Strong	Strong	Strong	Moderate
Jovanovic et al. (2014)	Strong	Strong	Moderate	Strong	Strong	Strong
Waters et al. (2014)	Strong	Strong	Moderate	Moderate	Moderate	Strong
Tzschoppe et al. (2014)	Strong	Moderate	Strong	Strong	Strong	Moderate
Den et al. (2015)	Strong	Moderate	Strong	Moderate	Moderate	Moderate
Haddad et al. (2015)	Strong	Strong	Strong	Strong	NA	Moderate
Kadosh et al. (2015)	Strong	Strong	Strong	Strong	NA	Strong
Shechner et al. (2015)	Strong	Strong	Strong	Strong	Strong	Strong
McGuire et al. (2016)	Strong	Moderate	Strong	Strong	Strong	Strong
Waters and Pine (2016)	Strong	Strong	Moderate	Moderate	Moderate	Strong
Schiele et al. (2016)	Strong	Moderate	Strong	Strong	NA	Moderate
Michalska et al. (2016)	Strong	Strong	Moderate	Strong	Strong	Moderate
McLaughlin et al. (2016)	Strong	Strong	Strong	Strong	Strong	Moderate
Newall et al. (2017)	Strong	Moderate	Weak	Strong	Strong	Strong
Jackson et al. (2016)	Strong	Moderate	Strong	Strong	NA	Moderate
Waters et al. (2017)	Strong	Strong	Moderate	Strong	Strong	Strong

Note. CS – Conditioned stimulus; US – Unconditioned stimulus; ACQ – Acquisition; EXT – Extinction.

Ratings were generally recorded via a computer keyboard, two studies used pen and pencil ratings and two studies stated that ratings were taken verbally. Absence of specific information was relatively common in the methods section in general but particularly with the specifics of CS evaluation measures.

Three studies measured subjective level of anxiety before conditioning, after conditioning and after extinction using a visual analogue scale all with 11 point scales from not at all anxious to very anxious, and typically assessed verbally by the experimenter. Haddad, Bilderbeck, James, and Lau (2015) used a nervousness scale during conditioning which asked participants "how nervous are you?" during each CS presentation, and a red bar appeared on the screen and they moved the bar with a button box along a 0–10 scale. It should be noted that some studies assessed anxiety with psychological assessments such as the Anxiety Diagnosis Inventory Scale (these measures were not examined in this review).

Three studies measured the intensity and unpleasantness of the US after the conditioning phase using a visual analogue scale on either 0–10 or 0–20 scale assessed verbally. One study asked participants to rate the scariness of the US after each phase on scale of 1–9 of not at all scary to very scary.

3.2.4. Contingency awareness

Sixteen studies measured contingency awareness, to ascertain whether participants identified which CS was associated with the US, this was completed verbally or with pen and paper. Studies varied in terms of timing of measuring contingency awareness after conditioning (8 studies) or after the extinction phase (2 studies). Six studies did not report when contingency awareness was assessed. Two studies inferred contingency awareness by whether participants correctly associated US expectancy with the CS+ and not the CS- on the last three trial of each CS during conditioning (Schiele et al., 2016; Waters, Theresiana, Neumann, & Craske, 2017).

Table 5 presents a summary of the results of differential fear conditioning and extinction for the healthy participants in all 35 studies. The table lists the age group of the participants (child or adolescent), a summarised description of the US type used and the dependent measures used. In Table 5, successful differential conditioning or extinction is indicated with a tick, and when unsuccessful this is indicated with a cross. Empty cells mean this measure was not reported. Successful extinction studies are shaded in groups based on the dependent measure used. Ten studies did not test extinction and are highlighted at the end of the table.

Successful differential conditioning was defined as there being a statistically significant difference in response magnitude to the CS + compared to the CS-. By defining differential conditioning in this way, it does not rule out that some degree of generalisation to the CS- may have occurred. Successful extinction was defined as no significant difference in response magnitude to the CS + compared to the CS- at the end of the extinction phase.[1] In 22 of 23 studies that used SCR, differential conditioning was successful. Using subjective reports (SR), 23 studies out of 25 had successful differential conditioning. Of the 9 studies that used FPS, 8 had successful differential conditioning.

[1] Five studies added a retest or reinstatement phase (Craske et al., 2008; Britton et al., 2013; Den et al., 2015; Michalska et al., 2016; and; Waters et al., 2017). All extinction phase results are recorded for the purpose of this review. Lau et al. (2008) conducted 3 trials of extinction on day 1 and 12 trials on day 2. Extinction results are recorded as per the study for day 2. Craske et al. (2008) conducted retest an average of 12 days after extinction, these results were consistent with the extinction results. Britton et al. (2013) and Michalska et al. (2016) conducted fMRI's three weeks after the extinction phase with morphed images to test recall. Den et al. (2015) and Waters et al. (2017) both tested reinstatement and retest of the same day as extinction with mixed results.

Table 4
Dependent measures used.

Authors	Data Collection Method	Units of Measures				CA
		SCR	SR	FPS	Other	
Block et al. (1970)	**Heart rate**	NA	NA	NA	Mean heart rate acceleration	×
Pliszka et al. (1993)	**SCR and heart rate**	SIR	NA	NA	Mean heart rate acceleration	CA
Liberman et al. (2006)	**SCR, SR and FPS**	SIR (Every second trial)	SAM 9 Point Likert (happiness, fear, perceived control and arousal) (phase)	WN (50 ms) 105 dB Footpoint to peak Magnitude	NA	×
Craske et al. (2008)	**SCR and SR**	AIR	SAM 5 point Likert (Valence and arousal) (Phase) paper and pencil	NA	NA	CA
Fairchild et al. (2008)	**SCR**	WIR	NA	NA	NA	CA
Lau et al. (2008)	**SR x2**	NA	Fear 10 point Likert (phase)	NA	NA	×
Neumann et al. (2008a)	**SCR and FPS**	SIR (Every second trial)	US Expectancy – dial and pointer (Trial-by-trial)	WN (50 ms) 110 dB Footpoint to peak Magnitude	NA	×
Neumann et al. (2008b)	**SCR and SR**	AIR	Paper and pencil SAM. 9-point Likert scale (Pleasantness and arousal) (phase) US Expectancy – dial and pointer (Trial-by-trial) US ratings (pleasantness, arousal and interest) (End of EXT)	NA	NA	×
Reeb-Sutherland et al. (2009)	**FPS**	NA	NA	WN (50 ms) 105 dB Baseline to peak	NA	×
Waters et al. (2009)	**SCR and SR**	AIR	SAM 5 point Likert (Valence, Arousal); Anxiety 11-point scale (phase) Verbal. US Ratings (Intensity and unpleasantness)	NA	NA	CA after ACQ
Fairchild et al. (2010)	**SCR**	WIR	NA	NA	NA	CA after EXT
Gao et al. (2015)	**SCR**	AIR	NA	NA	NA	×
Haddad et al. (2011)	**SR**	NA	9 point Likert scale (scariness and pleasantness). (phase)	NA	NA	×
Lau et al. (2011)	**SCR, SR and fMRI**	Only used in first experiment … no details	Online fear ratings (Trial-by-trial)	NA	BOLD contrast functional images	×
Glenn et al. (2012)	**FPS and SR**	NA	9-point scale (scary) Rated CS+, CS- and US (Phase)	WN (50 ms) 95 dB. Magnitudes	NA	CA
Patwell et al. (2012)	**SCR**	WIR	NA	NA	NA	×
Britton et al. (2013)	**SCR, SR, FPS and fMRI**	WIR (First interval only)	10 point Likert scale (fear) (phase)	AB (40 ms, 4–10 psi) Baseline to peak	3 weeks later BOLD contrast functional images	CA
Chauret et al. (2014)	**SCR and SR**	WIR (probe on foot)	5-point Likert scale using a button box (fear) (Trial-by-trial)	NA	NA	CA after EXT
Jovanovic et al. (2014)	**SCR and FPS**	WIR (First interval only)	US Exp end of each block and the end of EXT	WN (40 ms) 106 dB Peak amplitude	NA	×
Waters et al. (2014)	**SCR and SR**	AIR	SAM 9 point Likert (Valence and arousal) Anxiety level 11-point scale (phase) Verbal	NA	NA	CA after ACQ
Tzschoppe et al. (2014)	**SCR, SR and fMRI**	SIR	SAM 9 point Likert (Valence and arousal) Phase	NA	BOLD contrast functional images	CA after ACQ
Den et al. (2015)	**SR and Eye tracking**	NA	Scary ratings 8 point Likert (Phase) US Exp After ACQ and EXT and retest (phase)	NA	A diff score was calculated at each experimental phase to compare the latency of first fixation to the CS + relative to the CS-	
Haddad et al. (2015)	**fMRI and SR**	NA	Nervousness scale 10-point Likert (Trial-by-trial)	NA	BOLD contrast functional images	×
Kadosh et al. (2015)	**FPS**	NA	NA	AB 40 ms Baseline to peak amplitude	NA	CA after ACQ
Shechner et al. (2015)	**SCR, SR and FPS x2**	WIR (First interval only)	Fear level 10 point Likert (phase)		NA	×

(*continued on next page*)

Table 4 (*continued*)

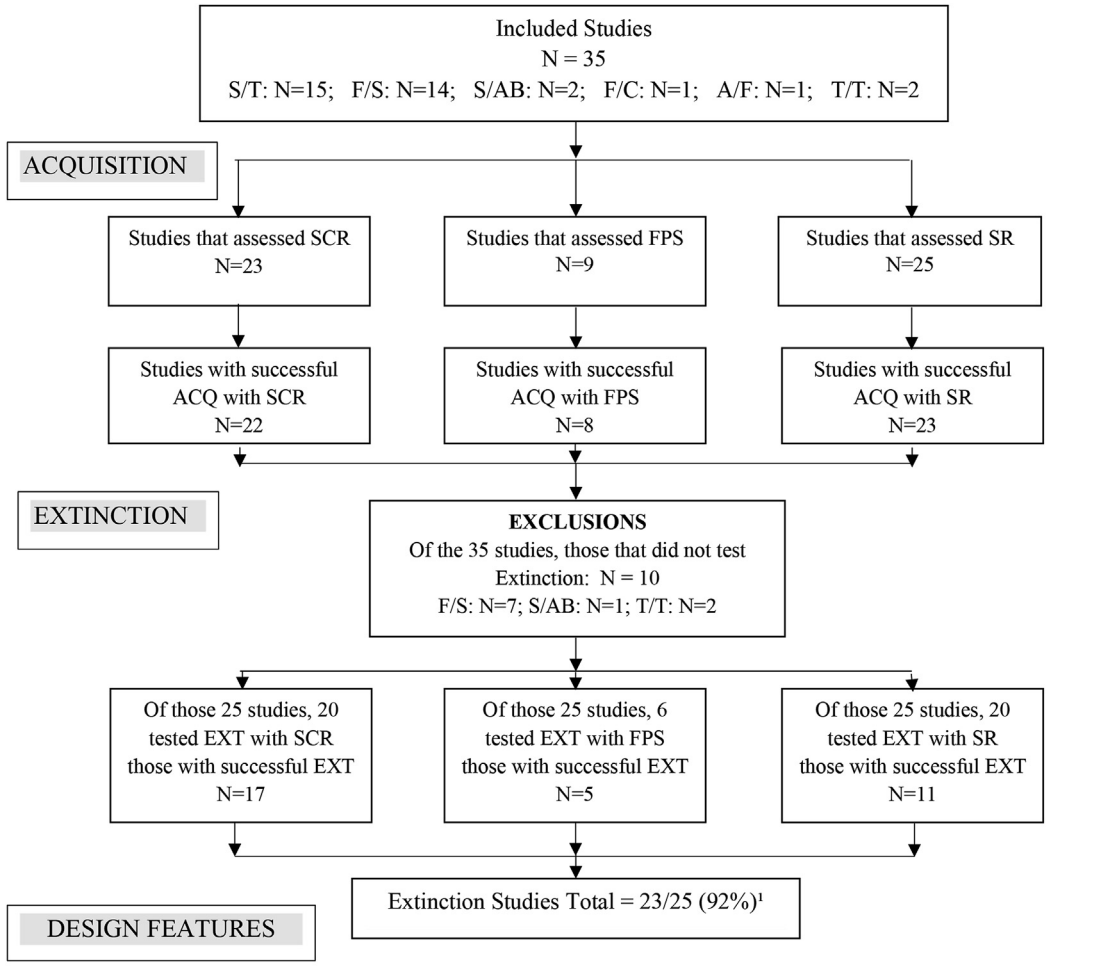
Authors	Data Collection Method	Units of Measures				CA
		SCR	SR	FPS	Other	
				AB (40 ms, 4–10 psi) Baseline to peak amplitude		
McGuire et al. (2016)	**SCR**	WIR	NA	NA	NA	×
Waters and Pine (2016)	**SCR and SR**	SIR	SAM 9 point Likert (valence and arousal) (Phase)	NA	NA	CA after ACQ
Schiele et al. (2016)	**SCR and SR**	WIR	9 point Likert scale (Arousal and valence) (Phase) US Exp 1–100% scale (Trial-by-Trial)	NA	NA	CA from US Exp results
Michalska et al. (2016)	**SCR and SR**	WIR	Fear level 10 point Likert (Phase)	NA	NA	×
McLaughlin et al. (2016)	**SCR and SR and fMRI**	SIR	Fear level 10 point Likert (Phase)	NA	Volume and thickness of amygdala and hippocampus	×
Newall et al. (2017)	**SR and NRT**	NA	Fear belief questionnaire 5 point Likert (Phase)	NA	Nature reserve task – measure behavioural avoidance	CA after ACQ
Jackson et al. (2016)	**FPS and SR**	NA	Fear level 5 point Likert (Phase)	WN (50 ms) Magnitude	NA	CA after ACQ
Waters et al. (2017)	**SR**	NA	Valence 11-point Likert scale (Trial-by-trial) Anxiety 11-point scale (Phase) US Exp (Trial-by-trial) US rating (Unpleasantness and intensity)	NA	NA	CA

Note: SCR – Skin conductance Response; SR – Subjective rating; FPS – Fear potentiated startle; CA – Contingency Awareness; US Exp – Unconditioned Stimulus Expectancy; SIR – Just 1st, 2nd or 3rd intervals (or a combination); AIR – Measuring all 1st, 2nd and 3rd intervals; WIR – Whole Interval response – could be Magnitude or Amplitude; WN – White noise; AB – Air Blast; NA – Not applicable.

Table 5
Successful Acquisition and Extinction Results Listed by Measure, US type and participant (child or adolescent).

Authors	Age	US Type	Successful Acquisition				Successful Extinction				Dropout Rate	Contingency Awareness results
			SCR	FPS	SR	Other	SCR	FPS	SR	Other		
Craske et al., 2008	Child	Pure Tone	✓		X		✓		X		0	67%-78% of each group aware
Waters et al., 2009	Child	Pure Tone	✓		✓		✓		X			100% of Anx reported correct CS-US relationship Only 55% of control did
Patwell et al., 2012	Child	White Noise	✓				✓					
Michalska et al., 2016	Child	Bell	✓		✓		✓		X		4	
Fairchild et al., 2010	Adolescent	White Noise	✓				✓					No differences
Fairchild et al., 2008	Adolescent	White noise	✓				✓					All groups aware
Chauret et al., 2014	Adolescent	Scream	✓		✓		✓		X			95% aware (3 p's unaware)
McGuire et al., 2016	Adolescent	Scream	✓				✓				7	
McLaughlin et al, 2016	Adolescent	Bell	✓		✓	✓	✓		X	X	2	
Shechner et al., 2015**	Adolescent	Scream	✓	✓	✓		✓	✓S=AB	X		3 – scream task	
		Bell	✓	✓	✓		✓	✓S=AB	X		3 – bell task	
Jovanovic et al., 2014*	Child	Airblast	✓	✓	✓		✓	✓S=WN	✓		10	
Neumann et al., 2008a	Adolescent	Scrape	✓	✓	✓		✓	✓S=WN	✓			
Britton et al., 2013*	Adolescent	Scream	✓	✓	✓		✓	✓S=AB	✓		7	100%. All p's reported that the CS+ screamed more than the CS-8 Anx and 13 control reported scream as unpredictable
Neumann et al., 2008b	Child	Scrape	✓		✓		✓		✓			
Waters et al., 2014	Child	Pure Tone	✓		✓		✓		✓			74% Con 57% Anx 83% Dep aware
Waters and Pine, 2016*	Child	Pure Tone	✓		✓		✓		✓			No sig difference between groups
Liberman et al., 2006	Child	Pure Tone	X	X	✓		X	X S=WN	✓		0	
Waters et al, 2017	Both	Scrape			✓				✓			100% aware – last 3 CS+ US Exp trials correct
Tzschoppe et al., 2014	Adolescent	Scream	✓		✓	✓	X		✓	✓		More are aware than unaware
Den et al., 2015	Adolescent	Scream			✓	✓			✓	✓	10	
Newall et al, 2017	Child	Fearful Face			✓	✓			✓	✓	Many excluded	Did not report findings
Pliszka et al., 1993**	Child	White noise	✓			✓	X			✓	0	100% aware
Lau et al., 2008**	Adolescent	Scream			✓				X		23 (ACQ)	
Haddad et al., 2011	Adolescent	Face and comments			✓				X			
Block et al., 1970	Child	Auto horn				✓					0	
Reeb-Sutherland et al., 2009	Adolescent	Airblast		✓								
Lau et al., 2011	Adolescent	Scream	✓		✓							
		Scream				✓						
Glenn, et al., 2012	Child	Scream		✓	X							80% aware
Haddad et al., 2015	Adolescent	Scream			✓	✓						
Kadosh et al., 2015	Adolescent	Scream		✓								The Control group were more aware than the Anx
Gao et al., 2015	Child	White noise	✓									
Schiele et al., 2016	Child	Scream	✓		✓						9	89% of adults and 48% of children aware
Jackson et al, 2017	Adolescent	Scream		✓	✓						11 - Fear	87% unaware

Note: AB = Airblast; WN = White noise; S = Stimulus; SCR – Skin conductance Response; SR – Subjective rating; FPS – Fear potentiated startle;
** Anxious patients result matched healthy controls
* Anxious patients had successful extinction on different dependent measures to healthy controls
The lower shaded panels are the studies that did not test extinction
Under the successful extinction columns, the shading highlights each dependent measures success.



Successful Extinction Studies N=23

Design Successful EXT/Total	S/T: N=15/15	F/S: N=6/7	S/AB N=1/1	A/F N=1/1
CS-US Reinforcement 100%	N=9	N=0	N=1	N=1
No of ACQ trials per CS >7	N=14	N=6	N=1	N=1
No of EXT trials per CS >5	N=10	N=6	N=1	N=1
Dependent measures Successful EXT/Total				
SCR	12/14	4/5	1/1	0/0
FPS	2/3	2/2	1/1	0/0
SR	5/11	3/6	1/1	1/1

Unsuccessful Extinction Studies N=2

Design Unsuccessful EXT/Total	F/S N=1/7	F/C N=1/1
CS-US Reinforcement 100%	N=0	N=0
No of ACQ trials per CS >7	N=1	N=1
No of EXT trials per CS >5	N=1	N=1
Dependent measures Successful EXT/Total		
SCR	0/0	0/0
FPS	0/0	0/0
SR	0/1	0/1

Fig. 2. Summary of integrated dependent measures and design features from the systematic review of 35 studies.
Note. SCR – Skin conductance Response; SR – Subjective rating; FPS – Fear potentiated startle; CS – Conditioned stimulus; US – Unconditioned stimulus; ACQ – Acquisition; EXT – Extinction.; S/T – Shape/Tone; F/S - Face/Scream; S/AB – Shape/Airblast; F/C – Face/Comment; A/F – Animal/Face; T/T – Tone/Tone
[1]Other dependent measures used include HR = 2; fMRI = 5; Other = 2. Not included in this review.

The studies in Table 5 were organised based on the successful extinction dependent measures and age group within each dependent measure. SCR was the most common measure used (23 studies in total), with 17 studies finding successful extinction, with 8 of the studies involving children and 9 involving adolescents. FPS was tested to extinction in six studies; four studies with adolescents and two with children, and of these, five demonstrated successful extinction. Finally, SR, which included CS evaluations and US Expectancy was successfully extinguished in 11 out of 20 studies that tested extinction (6 studies measured CS evaluations, 2 used US expectancy and 3 used a combination of US expectancy and CS evaluations). Nine extinction studies testing SR were conducted on children with one that tested both children and adolescents (Waters et al., 2017). Of those studies, six studies plus the Waters et al. (2017) study demonstrated successful extinction in SR measures.

Most reports of contingency awareness were descriptive. Five studies reported that all healthy participants were aware of the CS-US association, others reported that more participants were aware then

unaware; there was no difference between control and anxious; healthy participants reported greater awareness than anxious participants. The remaining studies reported the exact percentages of participants who demonstrated contingency awareness, with awareness ranged from 55% to 100% of participants (see Table 5).

Shechner et al. (2015) reported two studies comparing the use of different US types (scream and bell), and both revealed successful fear extinction in adolescents in FPS and SCR, but not SR (i.e., CS fear evaluations). FPS (airblast startle eliciting stimulus) and SCR was measured on every trial, and 6 adolescents dropped out due to fear in these studies. Liberman et al. (2006) also reported three dependent measures, however only found successful differential conditioning and extinction with CS evaluations. Three studies reported successful extinction observed in all three dependent measures. Britton et al. (2013) used face CSs with the face and scream US testing adolescent participants, measuring SCR, FPS using an air blast as the startle eliciting stimulus on every trial and measured the level of fear of the CS. They also reported a dropout rate during differential conditioning of 7 of the 42 healthy controls (16%). Jovanovic et al. (2014) used shape CSs with an air blast to the larynx as the US on child participants, measuring SCR, FPS using a white noise startle eliciting stimulus on every trial and measured US expectancy (by block and phase). They reported 10 participants dropping out due to fear (16%). Neumann et al. (2008) used shape CSs with the metal scraping on slate US in adolescent participants, measuring SCR, FPS using a white noise startle eliciting stimulus on half of the trials and measured US expectancy (trial-by-trial). No dropouts were reported in this study.

Table 5 also lists the proportion of participants who did not complete the study (Dropout Rate) as a result of their fear. The majority of participants reportedly dropped out during the conditioning phase. Of the 10 studies that reported dropouts due to fear, 5 studies used the scream US, 3 studies used the bell US, one used the air blast US and another the fearful face US. Also, it should be noted that in this table we have only reported the healthy controls who dropped out, not the anxious participants.

Sixteen studies included anxious patients as participants in their research (see Table 2). The extinction results from the patient groups matched the results of the healthy controls in 5/16 studies (31.25%); these studies are noted in Table 5 (**). In these five studies, three studies used the face/scream and two used shapes/tone as the CS and US respectively (Lau et al., 2008, did not find successful extinction in healthy and anxious participants). Three studies (marked with one * in Table 5) also found successful extinction with anxious patients, but on different dependent measures to the healthy controls. Review of these studies indicates considerable heterogeneity among the anxious samples in terms of diagnosis-type and clinical versus subclinical levels. This makes it more difficult to determine whether task parameters versus sample characteristics influence outcomes (see Supplementary Table 1, Extinction results for the non-control participants, for more details).

3.3. Design and experimental parameters

Fig. 2 provides a summary of the design features and measures employed across the reviewed studies. The upper part of the figure organises the studies by the type of CS/US combination and by the dependent measure used; SCR, FPS and SR (some overlap occurs due to multiple measures used within the same study). The number of studies achieving successful differential conditioning was 35, displayed on the third level by the dependent measure used (SCR N = 22; FPS N = 8; SR N = 22). Of those studies with successful differential conditioning, the middle panel of the figure displays studies which found successful extinction (those that did not test extinction are presented as exclusions). Successful extinction based on SCR occurred in 17 of the 20 studies that tested extinction with this measure, 5 of the 6 studies that tested extinction using FPS and 11 of 20 studies based on SR testing extinction using this measure (note that 21 studies overlap in measures used). Studies using other measures are also presented in notes, but not included due to small numbers (Pliszka et al., 1993; Tzschoppe et al., 2014; Den, Graham, Newall, & Richardson, 2015; Newell, Watson, Grant, & Richardson, 2017). The proportion of studies that achieved successful extinction in healthy youth controls were 23/25 (92%).

The lower panel of Fig. 2 summarises the design features of studies into two tables of successful extinction studies and unsuccessful extinction studies. The successful extinction table (left hand table) displays 100% of shape CS/tone US studies finding successful extinction (N = 15/15) and 85% of face CS/scream US studies finding successful extinction (N = 6/7). The design features table details the proportion of CS-US reinforcement schedule and number of trials used in these studies. A 100% reinforcement schedule was used in 60% of the shape CS/tone US studies whereas all 7 of the face CS/scream US studies used a schedule less than 100%. The majority of successful extinction studies had 7 or more trials per CS for conditioning and 5 or more trials per CS for extinction.

On the right hand table, the unsuccessful extinction studies are also listed with the design features detailed, one study used the face CS/scream US and one used the face CS/comments US combination (Haddad, Lissek, Pine, & Lau, 2011; Lau et al., 2008). These two studies both used a CS-US reinforcement schedule of less than 100% and 7 or more trials per CS for conditioning and 5 or more trials per CS for extinction.

The bottom rows of the two tables highlight the use of successful extinction/total of dependent measures by design. This highlights the variation of paradigms and measures throughout studies and the use of only one dependent measure used in the unsuccessful extinction studies.

4. Discussion

This review was conducted to compare the design and methodological features as well as the dependent measures of differential fear conditioning and extinction experiments with youth that revealed successful and unsuccessful extinction effects. Thirty-five differential fear conditioning and extinction studies with healthy youth between 2 and 17 years of age, published between 1970 and 2017, were identified. Of the 35 studies demonstrating successful differential conditioning, 10 studies did not test extinction and 2 studies did not report extinction in any measures. This resulted in 23 studies that documented successful extinction in at least one dependent measure. Extinction was found to be successful in 100% of studies using the shape CS/tone US combination and 85% in studies using the face CS/scream US combination. Magnitude of skin conductance responses was the most common dependent measure in which successful extinction was observed. Between phase subjective report measures were the second most frequently used dependent measure, CS evaluations were the most commonly used subjective rating, 9 out of 16 successful extinctions reported. Fear potentiated startle was the least commonly assessed dependent measure but found to be successful in 5 out of 6 extinction studies.

4.1. Design: experimental parameters

4.1.1. CS types

Shape CSs were used in 15 of the 23 successful extinction studies in healthy youth and 6 of the 23 successful extinction studies used face CSs. Shapes are a visual CS that are typically neutral prior to conditioning training and have been found to be less resistant to extinction than emotion evoking CS types such as faces (Lipp, 2006). Face CSs may evoke memories of aversive events in a participant's past or increase cognitive load due to the need for greater stimulus processing which may increase differential fear conditioning but may inhibit extinction (Lissek et al., 2005; Mineka & Oehlberg, 2008). Nevertheless, although shape CSs have been used almost twice as often as faces, the present

review found that the proportions of studies finding successful extinction was very similar for shape and face CSs (Shapes = 100%; Faces = 85%).

4.1.2. US types

A tone US was used in 15 of the 23 studies reporting successful extinction. A 1000 Hz pure tone was used in five of those 15 successful extinction studies and the decibels ranged from 87 to 107 dB (Craske et al., 2008; Liberman et al., 2006; Waters et al., 2009; Waters, Peters, Forrest, & Zimmer-Gembeck, 2014; Waters & Pine, 2016). The metal fork scraping on slate US delivered for 3 s at 83 dB was used in three of the 15 studies reporting successful extinction (Neumann, Waters, & Westbury, 2008; Neumann et al., 2008; Waters et al., 2017). An aversive bell sound delivered for 1 s at 95 dB was also used in a further three of the 15 studies finding successful extinction (McLaughlin et al., 2016; Michalska et al., 2016; Shechner et al., 2015). Furthermore, white noise was used in four of the 15 successful extinction studies (Patwell et al., 2012; Fairchild et al., 2008; Fairchild, Stobbe, van Goozen, Calder, & Goodyer, 2010 and Pliszka et al., 1993 – only tested HR).

Extinction was observed in 6 of 7 studies using the combination of a fearful face and scream as the US. The fearful face and scream combination was rated the most aversive and potent US, however the concern with youth participants is the high dropout rate (see Table 5). In the 7 studies that used faces and screams, high drop outs were recorded: 43% in Lau et al. (2008) and 45% in Britton et al. (2013), both healthy and patients dropped out due to anxiety invoked by the experiment. High dropout rates could suggest that using fearful faces and screams may result in a final sample that is problematic from an ethical perspective as it may result in a final sample that is biased or show different characteristics to the youth that have dropped out. Thus, tone stimuli of moderate intensity such as metal fork scraping on slate or aversive bell sounds appear to be capable of producing reliable differential conditioning effects to enable the observation of successful extinction effects, while also resulting in maximum likelihood of participation retention. Another suggestion could be trialling the scream US at a lower intensity to reduce the dropout of participants from fear.

4.1.3. Number of CS trials

Fourteen of the 15 successful extinction studies (93%) using shape CSs and tone USs included more than 7 conditioning trials per CS and 10/15 (66%) studies used more than 5 extinction trials per CS. All 6 successful extinction studies using face CSs and scream USs also included this number of trials per CS in conditioning and extinction. Across all studies, it was found that between 8 and 12 conditioning trials and between 6 and 12 extinction trials was the most successful combination for observing successful extinction effects. However, studies that have tended not to observe extinction effects on some measures such as CS evaluations have often included 8 trials and thus, it may be important for future studies to use 12 or more trials per CS as it may be the case that some dependent measures take longer to extinguish than others.

4.1.4. CS-US reinforcement schedule

It was predicted that 100% CS-US reinforcement schedule would be most successful for extinction studies. Out of the 23 successful extinction studies, only 11 used 100% reinforcement. Of the 15 successful extinction studies that used shape CSs and tone USs, nine (60%) of those used 100% reinforcement schedule of the US to the CS+ in the conditioning trials. None of the studies using face CSs and the scream USs used 100% reinforcement. However, the reinforcement schedule may be important depending on research goals. For example, 100% reinforcement would be recommended to ensure extinction of differential fear conditioning occurred, if reinstatement is of interest. For studies that aimed to test for fear generalisation rather than extinction, partial reinforcement may be preferable. Nevertheless, the studies reviewed here suggest that the CS-US reinforcement schedule did not appear to be a task parameter that affected whether or not extinction was observed. However, future studies should compare reinforcement schedules to confirm this empirically and determine whether the *rate of* extinction differs as a function of conditioning reinforcement schedule.

4.2. Design: dependent measures

4.2.1. Skin conductance responses

Skin conductance responses was the dependent measure with the most successful extinction, with 17 out of the 23 extinction studies.

In terms of measurement, as seen in Table 4, the most popular method of SCR measurement was using the whole interval response, with 48% of the 17 studies finding successful extinction utilising this approach. A recent investigation into the sensitivity of three different SCR indices, first interval responding (FIR), second interval responding (SIR) and entire interval responding found that more sensitive results come from designing and measuring conditioning experiments in such a way as to allow distinction between first and second interval responding (Luck & Lipp, 2016), which supports past recommendations (Prokasy & Kumpfer, 1973). The FIR is sensitive to orienting elicited by the CS onset whereas the SIR is sensitive to anticipation of the US and its absence. However, entire interval responding is not sensitive to these conceptually important distinctions between orienting and anticipation (Luck & Lipp, 2016).

4.2.2. Fear potentiated startle

Fear potentiated startle was assessed in 9 of the original 35 studies, with 8 of the 9 studies reporting successful differential conditioning. Of the 23 studies finding successful extinction, FPS was found to extinguish in 5 of 6 studies reporting FPS (Britton et al., 2013; Jovanovic et al., 2014; Neumann, Waters, & Westbury, 2008; Shechner et al., 2015). As identified, FPS has not been used extensively in youth studies, perhaps because the additional tones or air blasts required to elicit startle blinks may increase cognitive demands to determine which stimulus is associated with the CS+ and which are not, and the overall aversiveness of the task; which in turn, could affect learning or increase the dropout rate. Notably, of the studies that had the highest percentage of dropout, studies assessing FPS and the face CS and scream US were among them (Britton et al., 2013). However the authors of these studies generally attributed dropout to the scream US rather than the accumulated experience of the face CSs, the scream US and the tone or air blast to elicit FPS so this warrants further investigation (Britton et al., 2013; Jovanovic et al., 2014; Shechner et al., 2015).

Another consideration is that stimuli used to elicit startle blinks are likely to also affect SCRs. As seen in Table 4, variation also exists in whether FPS was indexed by blink magnitude versus amplitude and whether responses were assessed from either baseline to peak or foot point to peak. Another notable variation is the measurement of SCR whilst also measuring FPS, as the blink eliciting stimulus may interfere with the SCR. Variation includes only presenting the startle stimulus on half the trials and using the SCR results on the other half. Others have measured the FPS on every trial and the SCR at the very start of the interval response, just after CS onset. The meta-analysis by Lissek et al., 2005 recommended FPS as a dependent measure because it has good construct validity as a measure of fear and anxiety and also allows for the measurement of anxious arousal during ITIs that is not possible with SCR. Due to the limited number of studies using FPS in youth studies but the potential importance of this measure, more research is required to determine the most effective and tolerable experimental protocol using established guidelines for measuring and scoring FPS (Blumenthal et al., 2005).

4.3. Design: subjective report measures

4.3.1. Contingency awareness

Contingency awareness testing was included in 16 out of total 35

studies reviewed. Of the 23 successful extinction studies, 12 studies tested contingency awareness. Most studies did not report when contingency awareness was assessed, 8 reported it was after conditioning, and 2 reported it was after extinction.

Researchers have differed with respect to how to manage data from participants who are unaware of the CS+/US contingency. Although most researchers have included data from all participants including those without awareness (for example; Waters et al., 2014; Craske et al., 2008; Waters et al., 2009; Waters & Pine, 2016), on the basis that inclusion and exclusion of these participants did not alter results, one study reported excluding data from participants without contingency awareness (Schiele et al., 2016). This lack of difference may be due to a lack of power because most studies find only a small number of participants are unaware. To date, no studies have specifically examined factors that differentiate aware and unaware participants. However, Waters et al. (2017) found that participants who were less likely to report contingency awareness were younger (mean age: 7.8 years old) than those with awareness (mean age: 9.0 years old) and also, they may be more anxious (Waters et al., 2009).

4.3.2. US expectancy

Only six out of the 35 studies used the US expectancy measure, (only 5 out of the 23 successful extinction studies), which is a notable omission given findings from a recent study concluded that the US expectancy measure, particularly online (trial-by-trial) was a valuable and valid measurement for differential fear conditioning and extinction studies (Boddez, Baeyens, Luyten, & Vansteenwegen, 2013). Yet, providing trial-by-trial US expectancies can increase demands on participants and it is not known whether the requirement for children to make US expectancy ratings trial-by-trial differentially influences cognitive load and engagement, and thus, differential conditioning and extinction effects, relative to passive viewing of the CSs and obtaining between-phase ratings (Neumann et al., 2008). However, three studies that tested to extinction in this review used trial-by-trial US expectancy judgements and all reported successful extinction results (Neumann, Waters, & Westbury, 2008; Neumann et al., 2008; Waters et al., 2017). However, in Waters et al. (2017), children required more trials than adolescents and adults to establish differential conditioning and safety learning regarding the CS- assessed trial-by-trial. It is unclear if this reflects on cognitive load due to task demands of providing trial-by-trial ratings or genuine developmental differences in the rate of differential threat-safety learning. Of note, was that children reached similar levels of differential conditioning as adults by the end of the conditioning phase, so whether these differences in rate of learning play an important role in terms of differentially predicting fear and anxiety problems remains to be determined. Another notable difference between the successful studies is the variety in the scale of measurement, with only two studies using the same scale (see Table 4).

4.3.3. CS evaluations

Evaluations of the CSs were made in 22 of the 35 studies, of these 16 studies of the 23 successful extinction studies used CS evaluations, resulting in 9 studies with successful extinction results for CS evaluations. The variation amongst units and timing of measurement was wide-ranging (with scales including the self-assessment mannequin), and the most common being the use of Likert-type scales ranging from 5 to 11 points (9 being the most common). Fear, valence and arousal ratings of the CSs were the most common CS dimensions assessed. Valence evaluations have been assessed more often than the other CS dimensions, with negative valence of the CS + post-extinction strongly predicting return of fear in adults (Zbozinek, Hermans, Prenoveau, Liao, & Craske, 2015). Of the nine studies that tested CS valence evaluations, seven tested extinction. Waters et al. (2017) found that adolescents' negative CS + evaluations were more resistant to extinction and their positive evaluations of the CS- were more likely to decline during extinction relative to adults. This highlights the need to further investigate developmental differences between children and adolescents in CS evaluations and whether assessment method (i.e., trial by trial; between phase) influences outcomes.

4.3.4. Subjective anxiety ratings

Three of the 35 studies reviewed (Waters et al., 2009, 2014; Waters et al., 2017) reported assessing between phase subjective anxiety ratings. Two studies found a main effect of phase due to significantly lower anxiety ratings after extinction compared to before and after conditioning (Waters et al., 2009; 2014). The final study (Waters et al., 2017) also found that children and adolescent anxiety levels significantly declined from post-acquisition to post-extinction. Very few studies have included between phase anxiety ratings. However, given that they are an analogue measure of changes in anxiety following exposure therapy similar to subjective units of distress (SUDs) ratings, it would be informative for future studies to include these measures and to utilise a consistent range of responses such as an 11 point scale from not at all anxious to very anxious, verbally assessed.

4.3.5. Use of three dependent measures

Six of the 35 studies reviewed used all three dependent measures of SCR, FPS and SR to test differential fear conditioning and extinction (Britton et al., 2013; Jovanovic et al., 2014; Liberman et al., 2006; Neumann, Waters & Westbury, 2008; Shechner et al., 2015). Liberman et al. (2006) observed successful extinction effects among children only in SR (i.e., happiness, perceived control, fear and arousal) when tested between phase with a pure tone US and a white noise as the blink eliciting stimulus for FPS, occurring on every other trial, and SCR measured on the alternative trials; there was no drop out from this study. Shechner et al. (2015) only observed successful extinction among adolescents in two measures (SCR and FPS) using both faces/scream and shapes/tone CS/US combinations (Bell US) and the air blast as the blink eliciting stimuli, delivered 5 s after the CS onset to allow measurement of SCR in first 5 s of trial only; SR was tested between phase but did not extinguish. This study also had 18% of healthy participants dropout due to fear in both US type studies. Britton et al. (2013) successfully demonstrated extinction in adolescents in all three measures with the scream US and air blast as the blink eliciting stimulus, delivered every trial 5 s after CS onset; SCR was measured in the first 5 s of the trial only; SR of fear was tested between phase. However, the study had a dropout of 16% (46% when including the anxious adolescents), suggesting that although effective, the combination of the aversive scream US and the air blast blink eliciting stimulus was not well tolerated particularly amongst the anxious patients. Similarly, Jovanovic et al. (2014) found successful differential fear conditioning and extinction in FPS and SCR and US Expectancy (by phase) in children in which an air blast served as the US and white noise as the blink eliciting stimulus, delivered 6 s after CS onset in every trial and SCR measured from 3 to 6 s after CS onset. However, this study also had a relatively high drop out of 17% when including non-anxious and anxious participants. Neumann et al. (2008) reported successful extinction in adolescents using all three dependent measures, SCR, FPS and US Expectancy (trial-by-trial) with the metal scraping on slate US and white noise as the blink eliciting stimulus, delivered 7.5 s after CS onset on half of the trials and SCR measured on the alternative trials. This study reported no dropouts.

Interestingly, Neumann et al. (2008) and Liberman et al. (2006) each used the approach of assessing FPS and SCR on half the trials. This was to allow measurement of all intervals of the SCR on every second trial without interference from the startle response. Jovanovic et al. (2014), Britton et al. (2013) and Shechner et al. (2015) measured FPS on every trial, 5–6s after CS onset (500 ms–1000 ms before US) and measured SCR during the first few seconds, so as not to interfere with the startle response. However, this approach prevents the assessment of second and third interval SCRs, and with SCRs elicited in the first few seconds, they are unlikely to return to baseline by the time the startle

stimulus is delivered and this may affect the magnitudes of SCRs. Notably, all three of these studies had high dropout rates compared to no dropouts in the Neumann et al. (2008) and Liberman et al. (2006) studies, and although this is influenced by numerous factors, intolerance of electrodes to measure both SCR and FPS and experiencing both the aversive auditory US and startle stimulus may have been contributing factors.

4.4. Unsuccessful extinction studies

In establishing recommendations for design and methodological parameters and dependent measures derived from the literature, it is also helpful to consider the parameters of studies that did *not* find successful extinction. Only two studies that tested extinction were unsuccessful in all dependent measures that were assessed (i.e., Lau et al., 2008; Haddad et al., 2011) (see lower right hand table of Fig. 2). Both studies used adolescent participants and used faces as the CS type. Lau et al. (2008) used the female scream and fearful face as the US type at 75% CS-US reinforcement schedule, with the number of CS trials for differential conditioning and extinction at 16 and 15 respectively. The dependent measure was subjective reporting of fear ratings. Haddad et al. (2011) used face CSs and negative comments as the US at 75% CS-US reinforcement schedule, with the number of CS trials for differential conditioning and extinction at 9 and 8 respectively. The dependent measure was subjective ratings of the CS for scariness and pleasantness plus they used an attentional bias dot probe task. As these studies assessed SR dependent measures only, it is possible that inclusion of SCR and FPS measures might of yielded significant effects as observed in subsequent studies using the face CS and scream US combination (Britton et al., 2013; Chauret et al., 2014; Den et al., 2015; McGuire et al., 2016; Shechner et al., 2015; Tzschoppe et al., 2014). Also, in the Lau et al. (2008) the likelihood of observing extinction of subjective ratings may have been hampered by the large dropout of healthy and anxious participants (n = 23/54; 42%). As no other studies have used the face CS and negative comment US pairing in the Haddad et al. (2011) study, it is difficult to draw firm conclusions. Nevertheless, failure to observe extinction effects may have been due to the adolescent sample, the limited number of dependent measures assessed and sample retention problems.

4.5. Recommendations, unresolved issues and future directions

The present review highlights a number of key components of experimental paradigms and dependent measures that can be recommended for use in future research of differential fear conditioning and extinction studies in youth. These components are discussed independently based only on the studies in this review. The recommendations presented may vary if different combinations of these design components were utilised:

- Design – A differential fear conditioning and extinction paradigm is reliable[2]
- CS-US combinations – Both shape CSs with tone USs and face CSs with a scream US are effective, although risk of drop-out may need to be considered for the latter. It is recommended that shape CSs and the metal scrape on slate or alarm tone USs be used in studies involving children and either shape CSs and the scrape/alarm tone USs or face CSs and the scream US be used with adolescents
- Number of trials – Between 8 and 12 trials per CS is recommended to observe successful differential conditioning and extinction, preferably the higher end of this range if participant fatigue and engagement permit.
- Dependent measures – multiple measures are recommended including trial-by-trial measures of US expectancy, CS evaluation, SCR and FPS on alternative trials and between phase measures of CS valence, CS arousal and subjective anxiety.
- SCR should be measured using multiple latency onset intervals.
- Additional measures - Contingency awareness and US intensity and unpleasantness should be assessed after the conditioning phase.

The present study also highlights several key issues that remain unresolved and require attention in future research.

1. *Is the strength of differential conditioning and extinction different between shape CSs/tone USs vs face CSs/scream USs?* As seen in the results in Fig. 2, both combinations yielded successful differential conditioning and extinction. However, it is unclear if they differ in the *strength* of differential conditioning and extinction. Future studies comparing differential fear conditioning and extinction on multiple measures as a function of shapes/tones vs faces/scream would be useful.
2. *What is the most effective US to use in child and adolescent studies?* A study comparing differential fear conditioning and extinction in children and adolescents assigned to different US conditions (i.e., scream, pure tone, bells, scrape) would be useful. A prior study in which an electric shock stimulus, loud tone, and metal fork scraping on slate were compared in adults showed that the latter was as or more effective than the former two stimuli (Neumann & Waters, 2006).
3. *CS-US reinforcement schedule:* CS-US reinforcement schedule appeared not to vary systematically with studies finding successful extinction (see Fig. 2). However, further studies should examine whether the reinforcement schedule affects the rate of extinction and susceptibility to return of fear, particularly in combination with USs of differing intensity and aversiveness.
4. *What is the most effective experimental paradigm for assessing all three types of dependent measures?* The use of an aversive auditory blink eliciting stimulus and a mild auditory US such as pure tone may not be effective in producing differential conditioning and thus the examination of extinction effects (Liberman et al., 2006). However, a more aversive auditory US such as metal fork scraping on slate combined with a more aversive white noise blink eliciting stimulus may be suitable (Neumann et al., 2008). This combination has been tested in adolescents but not children (Neumann et al., 2008) and warrants examination. An aversive tactile air blast to the larynx is also aversive and has been used as the blink eliciting stimulus in several studies with the scream US and bell US only in adolescent participants (Britton et al., 2013; Shechner et al., 2015). However, in Jovanovic et al. (2014), they used air blast as the US with white noise as the blink eliciting stimulus on children finding successful extinction (in SCR, FPS and US expectancy) which highlights that air blast may in and of itself be an aversive stimulus. Although the delivery of one stimulus to the tactile modality and the other to the auditory modality may have the benefit of assisting children in discriminating between the stimuli and thus learning which stimulus is associated with the CS+ (Block et al., 1970), the use of air blast may be aversive for children and could affect retention. Two stimuli presented to the auditory modality appears to be the most effective solution. Further research is required that examines effects and tolerability of the metal fork scraping on slate US with the white noise blink eliciting stimulus with children and replicating the Jovanovic et al. (2014) study with children to see if they can tolerate the air blast US and the white noise blink eliciting stimulus in the same procedure.
5. *What is the most effective method for assessing CS evaluations and US expectancies?* Although obtaining multiple outcome measures is recommended, it remains unclear whether trial-by-trial compared to between phase assessment of US expectancies and CS evaluations

[2] Differential designs were the only designs reviewed in this paper due to no single cue designs at this stage being published for youth studies in humans.

differentially influences engagement and cognitive load during differential conditioning and extinction phases, particularly when combined with FPS as well as SCR measurement. As CS evaluations and US expectancy ratings are reliable and valid SR measures (Boddez et al., 2013) it will be important in future research to determine whether trial-by-trial versus between phase SR assessments impacts children's learning more so than adolescents and adults, and the implications of the timing of these measurements for assessing expectancies and evaluations during exposure therapy. As it would seem that multiple measures provides the greatest opportunity to elucidate differences in differential fear conditioning and extinction, the assessment of US expectancies, CS evaluations trial-by-trial and between phase is recommended. It is additionally recommended that all SR measures be assessed using a rating scale with the same range of response options to permit comparisons of outcomes across studies. Based on studies to date, this may include 1–9 ratings of CS valence and arousal using the SAM (Bradley & Lang, 1994) and 0–10 ratings of subjective anxiety assessed between phase and −5 to +5 on the keyboard for trial-by-trial US expectancy and CS evaluations (as many researchers may not have access to a dial and pointer device).

6. *Which measures should be considered primary and secondary outcome measures?* As observed in this review, most experiments observed extinction effects on at least one dependent measure but not on another. Drawing from the efficacy literature on treatment outcomes following exposure therapy (the clinical application of extinction) in which multiple outcome measures are also assessed, primary outcomes are often defined as clinician-rated, diagnostic measures and patient and parent symptom measures are treated as secondary measures (e.g., Spence et al., 2011). It may be informative for guiding future studies for researchers in the field to determine and define extinction based on primary and secondary outcome measures. Future studies that draw from both theoretical and empirical literatures on extinction outcomes are required to identify and test candidate primary and secondary outcome measures.

4.6. Limitations

The current review needs to be considered in relation to its limitations, such as the potential file drawer effect, i.e., there may be more research available that has not been published. Furthermore, researchers should consider that the individual variables considered in this review (e.g., CS type; number of trials) might interact in simple or complex ways and influence the extent of differential conditioning and extinction observed in a given study. For example, some CS-US combinations more so than others may require fewer pairings to achieve conditioning and more trials to achieve extinction, and these task parameters might produce different outcomes depending on the age of the participants. Recommendations cannot be provided about these interactions given that they have not been examined at the time of this review. However, this will be important to examine in future research.

We noted that the two unsuccessful extinction studies both included adolescents. We also noted that both of these studies included a limited range of dependent measures from which to observe extinction effects. It is noteworthy that a further 10 studies were not included in this review because they did not report extinction results. In some cases, it was stated in the manuscript that this was because differential fear conditioning was the only part of a differential fear conditioning and extinction experiment that was conducted (e.g., Glenn et al., 2012). However, in others, it was unclear whether extinction was conducted and findings were not reported or extinction was not assessed (e.g., Lau et al., 2011). Thus, although the present review was able to make some recommendations based on successful extinction studies, it was far more limited in being able to examine the features of studies that did not find successful extinction.

In this review “reliability” has been assessed across studies (i.e., how many studies observed reliable differential conditioning or extinction effects with certain procedures or measures). Other measures of reliability are also needed such as test-retest reliability with the same participants and procedures. It can be difficult to compare specific procedures or measures in terms of reliability and validity because the studies being compared often differ in other ways. This can confound conclusions regarding the more appropriate methods to use. This points to the need for more systematic investigations in which studies differ by only one methodological feature.

Finally, we developed a novel quality assessment instrument based on prior quality assessment protocols utilised in Effective Public Health Practice Project (EPHPP) (Thomas et al., 2004) to define and classify components of differential fear conditioning and extinction experiments with youth (see Supplementary file DFCE-QAI). Although preliminary and in need of further application and replication in reviews of other studies (e.g., adult studies and clinical populations), the reverse translation of standardised assessment and evaluation methodology used in clinical science research to experimental psychopathology research could assist with addressing problems concerning the reproducibility of findings and advancing the translation of experimental research on extinction into clinical science research and practice (see Craske, Hermans, & Vervliet, 2018; Tackett et al., 2017; Waters, Le Beau, & Craske, 2017).

5. Conclusion

This systematic review of differential fear conditioning and extinction studies in youth revealed several key features of the experimental designs and dependent measures of studies finding successful extinction results. The review identified that studies using differential conditioning and extinction paradigms with geometric shape CSs paired with an auditory US produced reliable extinction effects while maintaining good sample retention. The most common dependent measures were SCR and subjective ratings of CS fearfulness, valence and arousal. However, FPS is an effective but less commonly used dependent measure. Although further research is required to clarify various aspects of differential fear conditioning and extinction studies, it is recommended based on evidence at the time of this review that studies with children use shape CSs and an ecological tone USs such as the metal scrape on slate or the alarm bells; and studies with adolescents use either shape CSs and tone USs or face CSs with the scream US. Additionally, studies with both age groups should use 8 or more trials per CS in each phase and include (a) between phase CS valence and arousal ratings as well as subjective anxiety ratings and (b) trial-by-trial US expectancy ratings, CS evaluations, SCR and FPS, using a white noise startle eliciting stimulus on half the trials for each CS and assessing SCR from the non-startle trials. These recommendations have been formulated by applying rigorous qualitative assessment methods to the available studies at the time of this review. They are intended to inform the design and methods utilised in future research which in turn, may lead to further refinement of these recommendations over time.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2018.11.009.

References

Ashcroft, K. R., Guimarães, F. S., Wang, M., & Deakin, J. F. W. (1991). Evaluation of a psychophysiological model of classical fear-conditioning in anxious patients. *Psychopharmacology, 104*, 215–219.

Barry, T. J., Vervliet, B., & Hermans, D. (2016). Threat-related gaze fixation and its relationship with the speed and generalisability of extinction learning. *Australian Journal of Psychology, 68*, 200–208. https://doi.org/10.1111/ajpy.12124.

*Block, J. D., Sersen, E. A., & Wortis, J. (1970). Cardiac classical conditioning and reversal in the mongoloid, encephalopathic, and normal child. *Child Development,*

41(3), 771–785. https://doi.org/10.2307/1127223.
Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology, 42*(1), 1–15. https://doi.org/10.1111/j.1469-8986.2005.00271.x.
Boddez, Y., Baeyens, F., Luyten, L., & Vansteenwegen, D. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry, 44*(2), 201–206.
Boschen, M. J., Neumann, D. L., & Waters, A. M. (2009). Relapse of successfully treated anxiety and fear: Theoretical issues and recommendations for clinical practice. *Australian and New Zealand Journal of Psychiatry, 43*(2), 89–100. https://doi.org/10.1080/00048670802607154.
Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 49–59.
*Britton, J. C., Grillon, C., Lissek, S., Norcross, M. A., Szuhany, K. L., Chen, G., ... Pine, D. S. (2013). Response to learned threat: An fMRI study in adolescent and adult anxiety. *American Journal of Psychiatry, 170*(10), 1195–1204. https://doi.org/10.1176/appi.ajp.2013.12050651.
*Chauret, M., La Buissonnière-Ariza, V., Lamoureux Tremblay, V., Suffren, S., Servonnet, A., Pine, D. S., et al. (2014). The conditioning and extinction of fear in youths: what's sex got to do with it? *Biological Psychology, 100*, 97–105. https://doi.org/10.1016/j.biopsycho.2014.06.001.
Clum, G. A. (1969). A correlational analysis of the relationships between personality and perceptual variables and discriminant GSR conditioning. *Journal of Clinical Psychology, 25*(1), 33–35. https://doi.org/10.1002/1097-4679(196901)25:1<33::AID-JCLP2270250108>3.0.CO;2-K.
Colman, A. (2009). *A dictionary of Psychology*. Oxford University Press.
Craske, M. G., Hermans, D., & Vervliet, B. (2018). *State of the art and future directions for extinction as a translational model for fear and anxiety*, *Vol. 373*, Philosophical Transactions Royal Society Publishing20170025. https://doi.org/10.1098/rstb.2017.0025.
Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006.
*Craske, M. G., Waters, A. M., Bergman, R., Naliboff, B., Lipp, O. V., Negoro, H., et al. (2008). Is aversive learning a marker of risk for anxiety disorders in children? *Behaviour Research and Therapy, 46*(8), 954–967. https://doi.org/10.1016/j.brat.2008.04.011.
Dawson, M. E., Schell, A. M., & Courtney, C. G. (2011). The skin conductance response, anticipation, and decision-making. *Journal of Neuroscience, Psychology and Economics, 4*(2), 111–116.
Del-Ben, C. M., Vilela, J. A., Hetem, L. A., Guimarães, F. S., Graeff, F. G., & Zuardi, A. W. (2001). Do panic patients process unconditioned fear vs. conditioned anxiety differently than normal subjects? *Psychiatry Research, 104*(3), 227–237.
*Den, M. L., Graham, B. M., Newall, C., & Richardson, R. (2015). Teens that fear screams: A comparison of fear conditioning, extinction, and reinstatement in adolescents and adults. *Developmental Psychobiology, 57*(7), 818–832. https://doi.org/10.1002/dev.21330.
Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., ... Baas, J. M. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety, 32*(4), 239–253. https://doi.org/10.1002/da.22353.
Dunsmoor, J. E., Bandettini, P. A., & Knight, D. C. (2007). Impact of continuous versus intermittent CS-UCS pairing on human brain activation during Pavlovian fear conditioning. *Behavioral Neuroscience, 121*(4), 635–642.
*Fairchild, G., Stobbe, Y., van Goozen, S. H., Calder, A. J., & Goodyer, I. M. (2010). Facial expression recognition, fear conditioning, and startle modulation in female subjects with conduct disorder. *Biological Psychiatry, 68*(3), 272–279. https://doi.org/10.1016/j.biopsych.2010.02.019.
*Fairchild, G., Van Goozen, S. H., Stollery, S. J., & Goodyer, I. M. (2008). Fear conditioning and affective modulation of the startle reflex in male adolescents with early-onset or adolescence-onset conduct disorder and healthy control subjects. *Biological Psychiatry, 63*(3), 279–285.
Fayu, C. (1961). Fear conditioning with normals, neurotics, & schizophrenics. *Acta Psychologica Taiwanica, 3*, 18–33.
Gao, Y., Raine, A., Venables, P. H., Dawson, M. E., & Mednick, S. A. (2010a). Association of poor childhood fear conditioning and adult crime. *American Journal of Psychiatry, 167*(1), 56–60. https://doi.org/10.1176/appi.ajp.2009.09040499.
Gao, Y., Raine, A., Venables, P. H., Dawson, M. E., & Mednick, S. A. (2010b). 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.
Gao, Y., Raine, A., Venables, P. H., Dawson, M. E., & Mednick, S. A. (2010c). Reduced electrodermal fear conditioning from ages 3 to 8 years is associated with aggressive behavior at age 8 years. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 51*(5), 550–558. https://doi.org/10.1111/j.1469-7610.2009.02176.x.
*Gao, Y., Tuvblad, C., Schell, A., Baker, L., & Raine, A. (2015). Skin conductance fear conditioning impairments and aggression: A longitudinal study. *Psychophysiology, 52*(2), 288–295. https://doi.org/10.1111/psyp.12322.
Ginsberg, G. S., Becker, E. M., Keeton, C. P., Sakolsky, D., Piacentini, J., Albano, A. M., ... Kendall, P. C. (2014). Naturalistic follow-up of youths treated for pediatric anxiety disorders. *JAMA Psychiatry, 71*(3), 310–318. https://doi.org/10.1001/jamapsychiatry.2013.4186.
*Glenn, C. R., Klein, D. N., Lissek, S., Britton, J. C., Pine, D. S., & Hajcak, G. (2012). The development of fear learning and generalization in 8–13 year-olds. *Developmental Psychobiology, 54*(7), 675–684. https://doi.org/10.1002/dev.20616.
Grillon, C. (2002). Startle reactivity and anxiety disorders: Aversive conditioning, context, and neurobiology. *Biological Psychiatry, 52*(10), 958–975. https://doi.org/10.1016/S0006-3223(02)01665-7.
Grillon, C., Merikangas, K. R., Dierker, L., Snidman, N., Arriaga, R. I., Kagan, J., et al. (1999). Startle potentiation by threat of aversive stimuli and darkness in adolescents: A multi-site study. *International Journal of Psychophysiology, 32*, 63–73.
*Haddad, A. D., Bilderbeck, A., James, A. C., & Lau, J. Y. (2015). Fear responses to safety cues in anxious adolescents: Preliminary evidence for atypical age-associated trajectories of functional neural circuits. *Journal of Psychiatric Research, 68*, 301–308. https://doi.org/10.1016/j.jpsychires.2015.05.009.
*Haddad, A. D., Lissek, S., Pine, D. S., & Lau, J. Y. (2011). How do social fears in adolescence develop? Fear conditioning shapes attention orienting to social threat cues. *Cognition & Emotion, 25*(6), 1139–1147. https://doi.org/10.1080/02699931.2010.524193.
Halberstam, J. L. P. D. (1961). Some personality correlates of conditioning, generalization, and extinction. *Psychosomatic Medicine, 23*(1), 67–76.
Hermann, C., Ziegler, S., Birbaumer, N., & Flor, H. (2002). Psychophysiological and subjective indicators of aversive Pavlovian conditioning in generalized social phobia. *Biological Psychiatry, 52*(4), 328–337. https://doi.org/10.1016/S0006-3223(02)01385-9.
Higa-McMillan, C. K., Francis, S. E., Rith-Najarian, L., & Chorpita, B. F. (2016). Evidence base update: 50 Years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child and Adolescent Psychology, 45*(2), 91–113. https://doi.org/10.1080/15374416.2015.1046177.
Howe, E. S. (1957). GSR conditioning in anxiety states, normal and chronic functional schizophrenic subjects. *Journal of Abnormal and Social Psychology, 56*, 183–189. https://doi.org/10.1037/h0047365.
*Jackson, F., Nelson, B. D., Meyer, A., & Hajcak, G. (2016). Pubertal development and anxiety risk independently relate to startle habituation during fear conditioning in 8-14 year-old females. *Developmental Psychobiology, 59*, 436–448. https://doi.org/10.1002/dev.21506.
*Jovanovic, T., Nylocks, K. M., Gamwell, K. L., Smith, A., Davis, T. A., Norrholm, S. D., et al. (2014). Development of fear acquisition and extinction in children: Effects of age and anxiety. *Neurobiology of Learning and Memory, 113*, 135–142. https://doi.org/10.1016/j.nlm.2013.10.016.
*Kadosh, K. C., Haddad, A. D. M., Heathcote, L. C., Murphy, R. A., Pine, D. S., & Lau, J. Y. F. (2015). High trait anxiety during adolescence interferes with discriminatory context learning. *Neurobiology of Learning and Memory, 123*, 50–57. https://doi.org/10.1016/j.nlm.2015.05.002.
Kendall, P. C., Robin, J. A., Hedtke, K. A., Suveg, C., Flannery-Schroeder, E., & Gosch, E. (2005). Considering CBT with anxious youth? Think exposures. *Cognitive and Behavioral Practice, 12*, 136–150.
Lang, P. J. (1980). Behavioral treatment and bio-behaviorial assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.). *Technology in mental health care delivery systems* (pp. 119–137). Norwood, NJ: Ablex.
*Lau, J. Y., Britton, J. C., Nelson, E. E., Angold, A., Ernst, M., Goldwin, M., ... Pine, D. S. (2011). Distinct neural signatures of threat learning in adolescents and adults. *Proceedings of the National Academy of Sciences of the United States of America, 108*(11), 4500–4505. https://doi.org/10.1073/pnas.1005494108.
*Lau, J. Y., Lissek, S., Nelson, E. E., Lee, Y., Roberson-Nay, R., Poeth, K., ... Pine, D. S. (2008). Fear conditioning in adolescents with anxiety disorders: Results from a novel experimental paradigm. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(1), 94–102. https://doi.org/10.1097/chi.0b01e31815a5f01.
*Liberman, L. C., Lipp, O. V., Spence, S. H., & March, S. (2006). Evidence for retarded extinction of aversive learning in anxious children. *Behaviour Research and Therapy, 44*(10), 1491–1502. https://doi.org/10.1016/j.brat.2005.11.004.
Lipp, O. V. (2006). Human fear learning: Contemporary procedures and measurement. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.). *Fear and learning: From basic processes to clinical implications* (pp. 37–51). Washington, DC, US: American Psychological Association.
Lissek, S., Baas, J. M. P., Pine, D. S., Orme, K., Dvir, S., Nugent, M., ... Grillon, C. (2005). Airpuff startle probes: An efficacious and less aversive alternative to white-noise. *Biological Psychology, 68*(3), 283–297. https://doi.org/10.1016/j.biopsycho.2004.07.007.
Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., et al. (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behaviour Research and Therapy, 43*(11), 1391–1424. https://doi.org/10.1016/j.brat.2004.10.007.
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.
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.
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. https://doi.org/10.1016/j.jesp.2009.08.001.
*McGuire, J. F., Orr, S. P., Wu, M. S., Lewin, A. B., Small, B. J., Phares, V., ... Storch, E. A. (2016). Fear conditioning and extinction in youth with obsessive-compulsive disorder. *Depression and Anxiety, 33*, 229–237. https://doi.org/10.1002/da.22468.
*McLaughlin, K. A., Sheridan, M. A., Gold, A. L., Duys, A., Lambert, H. K., Peverill, M., ... Pine, D. S. (2016). Maltreatment exposure, brain structure and fear conditioning in children and adolescents. *Neuropsychopharmacology, 41*, 1956–1964. https://doi.org/10.1038/npp.2015.365.
*Michalska, K. J., Shechner, T., Hong, M., Britton, J. C., Leibenluft, E., Pine, D. S., et al. (2016). A developmental analysis of threat/safety learning and extinction recall

during middle childhood. *Journal of Experimental Child Psychology, 146*, 95–105. https://doi.org/10.1016/j.jecp.2016.01.008.

Mineka, S., & Oehlberg, K. (2008). The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. *Acta Psychologica, 127*(3), 567–580. https://doi.org/10.1016/j.actpsy.2007.11.007.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Monk, C. S., Grillon, C., Baas, J. M. P., McClure, E. B., Nelson, E. E., Zarahn, E., ... Pine, D. S. (2003). A neuroimaging method for the study of threat in adolescents. *Developmental Psychobiology, 43*, 359–366. https://doi.org/10.1002/dev.10146.

Neumann, D. L., & Waters, A. M. (2006). The use of an unpleasant sound as an unconditional stimulus in a human aversive Pavlovian conditioning procedure. *Biological Psychology, 73*, 175–185. https://doi.org/10.1016/j.biopsycho.2006.03.004.

*Neumann, D. L., Waters, A. M., & Westbury, H. (2008a). The use of an unpleasant sound as the unconditional stimulus in aversive Pavlovian conditioning experiments that involve children and adolescent participants. *Behavior Research Methods, 40*, 622–625. https://doi.org/10.3758/BRM.40.2.622.

*Neumann, D. L., Waters, A. M., Westbury, H., & Henry, J. (2008b). The use of an unpleasant sound unconditional stimulus in an aversive conditioning procedure with 8- to 11-year-old children. *Biological Psychology, 79*, 337–342. https://doi.org/10.1016/j.biopsycho.2008.08.005.

*Newall, C., Watson, T., Grant, K.-A., & Richardson, R. (2017). The relative effectiveness of extinction and counter-conditioning in diminishing children's fear. *Behaviour Research and Therapy, 95*, 42–49. https://doi.org/10.1016/j.brat.2017.05.006.

Norrholm, S. D., Jovanovic, T., Olin, I. W., Sands, L. A., Karapanou, I., Bradley, B., et al. (2011). Fear extinction in traumatized civilians with posttraumatic stress disorder: Relation to symptom severity. *Society of Biological Psychiatry, 69*, 556–563.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483.

*Pattwell, S. S., Duhoux, S., Hartley, C. A., Johnson, D. C., Jing, D., Elliott, M. D., ... Lee, F. S. (2012). Altered fear learning across development in both mouse and human. *Proceedings of the National Academy of Sciences of the United States of America, 109*(40), 16318–16323. https://doi.org/10.1073/pnas.1206834109.

Peri, T., Ben-Shakhar, G., Orr, S. P., & Shalev, A. Y. (2000). Psychophysiologic assessment of aversive conditioning in posttraumatic stress disorder. *Biological Psychiatry, 47*(6), 512–519. https://doi.org/10.1016/S0006-3223(99)00144-4.

Pine, D., Fyer, A., Grun, J., Phelps, E. A., Szeszko, P. R., Koda, V., ... Bilder, R. M. (2001). Methods for developmental studies of fear conditioning circuitry. *Society of Biological Psychiatry, 50*, 225–228.

*Pliszka, S. R., Hatch, J. P., Borcherding, S. H., & Rogeness, G. A. (1993). Classical conditioning in children with attention deficit hyperactivity disorder (ADHD) and anxiety disorders: A test of quay's model. *Journal of Abnormal Child Psychology, 21*(4), 411–423.

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). USA: Academic Press.

Rapee, R. M., Schniering, C. A., & Hudson, J. L. (2009). Anxiety disorders during childhood and adolescence: Origins and treatment. *Annual Review of Clinical Psychology, 5*, 311–341.

*Reeb-Sutherland, B. C., Helfinstein, S. M., Degnan, K. A., Perez-Edgar, K., Henderson, H. A., Lissek, S., ... Fox, N. A. (2009). Startle response in behaviorally inhibited adolescents with a lifetime occurrence of anxiety disorders. *Journal of American Academy of Child and Adolescence Psychiatry, 48*(6), 610–617. https://doi.org/10.1097/CHI.0b013e31819f70fb.

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. https://doi.org/10.1111/j.1469-8986.2011.01308.x.

*Schiele, M. A., Reinhard, J., Reif, A., Domschke, K., Romanos, M., Deckert, J., et al. (2016). Developmental aspects of fear: Comparing the acquisition and generalization of conditioned fear in children and adults. *Developmental Psychobiology, 58*(4), 471–481. https://doi.org/10.1002/dev.21393.

Schneider, F., Weiss, U., Kessler, C., Muller-Gartner, H.-W., Posse, S., Salloum, J. B., ... Birbaumer, N. (1999). Subcortical correlates of differential classical conditioning of aversive emotional reactions in social phobia. *Society of Biological Psychiatry, 45*, 863–871.

Sehlmeyer, C., Schoning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., et al. (2009). Human fear conditioning and extinction in neuroimaging: A systematic review. *PloS One, 4*(6), e5865. https://doi.org/10.1371/journal.pone.0005865.

*Shechner, T., Britton, J. C., Ronkin, E. G., Jarcho, J. M., Mash, J. A., Michalska, K. J., ... Pine, D. S. (2015). Fear conditioning and extinction in anxious and nonanxious youth and adults: Examining a novel developmentally appropriate fear-conditioning task. *Depression and Anxiety, 32*(4), 277–288. https://doi.org/10.1002/da.22318.

Shechner, T., Hong, M., Britton, J. C., Pine, D. S., & Fox, N. A. (2014). Fear conditioning and extinction across development: Evidence from human studies and animal models. *Biological Psychology, 100*, 1–12. https://doi.org/10.1016/j.biopsycho.2014.04.001.

Spence, S. H., Donovan, C. L., March, S., Gamble, A., Anderson, R. E., Prosser, S., et al. (2011). A randomized controlled trial on online versus clinic-based CBT for adolescent anxiety. *Journal of Consulting and Clinical Psychology, 79*, 629–642. https://doi.org/10.1037/a0024512.

Steinberg, L. (2011). *Adolescence.* New York, NY: McGraw-Hill.

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., et al. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science, 12*, 742–756. https://doi.org/10.1177/1745691617690042.

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public Health nursing interventions. *Worldviews on Evidence-based Nursing, 1*(3), 176–184. https://doi.org/10.1111/j.1524-475X.2004.04006.x.

*Tzschoppe, J., Nees, F., Banaschewski, T., Barker, G. J., Büchel, C., Conrod, P. J., ... Flor, H. (2014). Aversive learning in adolescents: Modulation by amygdala-prefrontal and amygdala-hippocampal connectivity and neuroticism. *Neuropsychopharmacology, 39*(4), 875–884. https://doi.org/10.1038/npp.2013.287.

Van den Bergh, O., Stegen, K., & Van de Woestijne, K. P. (1997). Learning to have psychosomatic complaints. *Psychosomatic Medicine, 59*(1), 13–23.

Waters, A. M., & Craske, M. G. (2016). Towards a cognitive-learning formulation of youth anxiety: A narrative review of theory and evidence and implications for treatment. *Clinical Psychology Review, 50*, 50–66.

*Waters, A. M., Henry, J., & Neumann, D. L. (2009). Aversive pavlovian conditioning in childhood anxiety disorders: Impaired response inhibition and resistance to extinction. *Journal of Abnormal Psychology, 118*(2), 311–321. https://doi.org/10.1037/a0015635.

Waters, A. M., LeBeau, R., & Craske, M. G. (2017A). Experimental psychopathology and clinical psychology: An integrative model to guide clinical science and practice. *Psychopathology Review* (in press).

*Waters, A. M., Peters, R.-M., Forrest, K. E., & Zimmer-Gembeck, M. (2014). Fear acquisition and extinction in offspring of mothers with anxiety and depressive disorders. *Developmental Cognitive Neuroscience, 7*, 30–42. https://doi.org/10.1016/j.dcn.2013.10.007.

*Waters, A. M., & Pine, D. S. (2016). Evaluating differences in Pavlovian fear acquisition and extinction as predictors of outcome from cognitive behavioural therapy for anxious children. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1111/jcpp.12522.

*Waters, A. M., Theresiana, C., Neumann, D. L., & Craske, M. G. (2017). Developmental differences in aversive conditioning, extinction, and reinstatement: A study with children, adolescents, and adults. *Journal of Experimental Child Psychology, 159*, 263–278. http://doi.org/10.1016/j.jecp.2017.02.012.

Williams, L. M., Barton, M. J., Kemp, A. H., Liddell, B. J., Peduto, A., Gordon, E., et al. (2005). Distinct amygdala–autonomic arousal profiles in response to fear signals in healthy males and females. *NeuroImage, 28*(3), 618–626. https://doi.org/10.1016/j.neuroimage.2005.06.035.

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 & Emotion, 29*(4), 654–667. https://doi.org/10.1080/02699931.2014.930421.