HHS Public Access

Author manuscript

Behav Res Ther. Author manuscript; available in PMC 2021 February 01.

Published in final edited form as:

Behav Res Ther. 2020 February ; 125: 103532. doi:10.1016/j.brat.2019.103532.

Behavioral and neural processes in counterconditioning: past and future directions

Nicole Keller[1], **Augustin Hennings**[1], **Joseph Dunsmoor**[1,2]

[1]University of Texas at Austin, Institute for Neuroscience, Austin, TX, USA, 78712

[2]University of Texas at Austin, Department of Psychiatry, Austin, TX, USA, 78712

Abstract

Counterconditioning refers both to the technique and putative process by which behavior is modified through a new association with a stimulus of an opposite valence. Similar to extinction, counterconditioning is considered a form of inhibition that interferes with the expression of the originally learned response without erasing it. But whereas interest in extinction continues to rise, counterconditioning has received far less attention. Here, we provide an in-depth review of counterconditioning research and detail whether counterconditioning is any more effective than extinction at preventing relapse of the originally learned behavior. We consider the clinical implications of counterconditioning, describe recent neurobiological and neuroimaging research in this area, and consider future avenues in need of further investigation.

Keywords

associative learning; extinction; counterconditioning; Pavlovian conditioning; amygdala; aversive-appetitive interactions; systematic desensitization

For over a century psychologists have investigated how to effectively and persistently eliminate maladaptive behavior and unwanted memories. The clinical relevance of this knowledge is straightforward, as symptoms of many mental health disorders can be characterized to some degree as abnormalities of learning, memory, and behavior (Ressler & Mayberg, 2007). This includes intrusive memories (Trauma and Stress-related Disorders), addiction (Substance Abuse Disorder), excessive worrying (Generalized Anxiety Disorder), compulsive thoughts or behaviors (Obsessive Compulsive Disorder), and fear and avoidance of particular stimuli, people, or situations (Phobias, Panic Disorder, Trauma). Thus, there is a strong motivation to translate advances in research on learning and memory to provide insight into the neurobiological mechanisms of mental health disorders and innovate clinical treatment. But for all the advances in the psychology and neuroscience of learning and memory, techniques to persistently modify maladaptive behavior in humans remain elusive

*Correspondence to: joseph.dunsmoor@austin.utexas.edu.

(M. E. Bouton, 2014; LeDoux & Pine, 2016). Theoretical and laboratory research in behavior modification is largely encompassed within the field of associative learning, with a particular focus on the extinction of conditioned behavior. Here, we review another time-honored form of behavior modification, counterconditioning.

The term counterconditioning refers both to the technique and the putative process by which behavior is modified through a new association with a stimulus of an opposite valence (Figure 1). Counterconditioning holds particular relevance in the history of psychological research due to its direct influence on experimentally informed therapies for mental health (Joseph Wolpe & Plaud, 1997); in particular, it was central to the development of behavioral therapy techniques such as systematic desensitization (Joseph Wolpe, 1954). The concept of opponent appetitive-aversive interactions has also served an important role in theories of motivation and emotion (Solomon & Corbit, 1978) and neurobiological models of affective value coding in the brain (Berridge, 2019; Daw, Kakade, & Dayan, 2002; Grossberg, 2000). But whereas interest in the psychological and neural mechanisms of experimental extinction continues to rise, interest in counterconditioning has largely held steady over the last several decades.

In this review, we integrate human and rodent research on counterconditioning with contemporary learning and memory accounts of inhibiting and overriding maladaptive behaviors and memories. A prevailing issue throughout this review concerns the longstanding question of whether counterconditioning and extinction are distinct processes, or whether counterconditioning is merely a technique to augment the same processes involved in extinction. Early learning theories built on the concept that motivated behavior is maintained by the distinctly reinforcing properties of aversive or appetitive stimuli (Hull, 1943; J. Konorski, 1967; Pavlov, 1927; B. F. Skinner, 1938). Interactions between aversive and appetitive systems were assumed by many early theories to reciprocally inhibit one another to guide behavior, such that activation of the aversive system inhibits appetitive processes, and vice versa (Dickinson & Pearce, 1977; Robert A Rescorla & Solomon, 1967). Counterconditioning experiments provide evidence that an unconditioned stimulus (US) of the opposite valence can hasten the reduction of the original conditioned response (CR) and give rise to a new set of behaviors consistent with the new US (Dickinson & Pearce, 1977; John M Pearce & Dickinson, 1975). But, on the other hand, there a reported instances where counterconditioning techniques are no more effective (or even less effective) at modifying behavior than simply omitting the US (e.g., Holmes, Leung, & Westbrook, 2016). And like extinguished behaviors, counter-conditioned behaviors are prone to revert to the originally learned response when later tested in isolation (M. E. Bouton, 2004). These types of findings, detailed below, cast doubt on a special process of counterconditioning, *per se*, and perhaps indicate that counterconditioning is an approach to (sometimes) bolster a conserved extinction process (M. E. Bouton & Peck, 1992; Lomont, 1965).

We also detail recent neuroscience research using counterconditioning approaches that reveal neural processes involved in counteracting the return of the originally learned behavior, including defensive and drug seeking behavior. We also discuss an area of human research where counterconditioning has consistently shown to be more effective than standard extinction; that is, changing learned preferences (Kerkhof, Vansteenwegen,

Baeyens, & Hermans, 2011; Ludvik, Boschen, & Neumann, 2015). And while human neuroimaging of counterconditioning is limited, we describe emerging approaches integrating counterconditioning and real-time neurofeedback, which effectively bypasses the need to confront highly emotional stimuli in order to change their neural representations.

A short history of counterconditioning

As the term implies, counterconditioning is derived from research in Pavlovian and operant conditioning. In a typical Pavlovian counterconditioning design, a neutral conditioned stimulus (CS, e.g., a tone) is first paired with a biologically salient US of a particular valence (e.g., shock or food) and then paired with a US of the opposite valence (e.g., food or shock, respectively). In rodents, performance at test is oftentimes measured as the amount of time either freezing to the CS, the amount of head-jerks to the CS or head-entries to the location of food delivery (i.e., magazine entries). Operant conditioning designs, in contrast, tend to measure performance on active avoidance or conditioned suppression tasks. In both Pavlovian and operant counterconditioning preparations, the CS begins to elicit a CR appropriate to the new US (i.e., US2). For clinical purposes, the relevance of counterconditioning is to change maladaptive associations between cues associated with outcomes of a particular value so that, for example, feared objects or avoidant behaviors are associated with safety (such as desensitization) or pleasurable cues or maladaptive approach behaviors are regarded as aversive (such as aversion therapy).

The earliest experimental report of counterconditioning in the human literature was the “Little Peter” experiment by Mary Cover Jones (1924). The experiment followed the infamous “Little Albert” experiment by Watson and Rayner (1920), in which Albert, an 11-month infant, was conditioned to fear a white rat by presentations of a frighteningly loud noise when he reached for the rat. In the Cover Jones experiment, a child named Peter, who was nearly 3 years old at the start of the experiment, showed signs of fear towards certain stimuli, including a white rabbit. However, by placing the rabbit in the room near Peter while he ate candy (an appetitive stimulus), his fear of the rabbit gradually subsided. Over a number of sessions, the feared rabbit was moved closer and closer to Peter. Eventually, Peter allowed the rabbit to nibble his fingers without fear. While not quite as notorious as “Little Albert,” the “Little Peter” experiment is considered one of the earliest laboratory demonstrations of behavior modification, and a precursor to systematic desensitization therapies that are still a widely used (Tryon, 2005; J. Wolpe, 1961).

Central to the clinical account of counterconditioning is that behaviors associated with appetitive systems inhibit behaviors associated with aversive systems, and vice-versa, referred to as *reciprocal inhibition*. Reciprocal inhibition therapies (Joseph Wolpe, 1954) therefore invoke feelings or actions that are presumably incompatible with the unwanted reactions, such as a patient being in a fully relaxed state in advance of a tense anxiety provoking situation. An important principle of reciprocal inhibition therapy is that the stimulus eliciting the competing response must be stronger than the stimulus eliciting the unwanted response. For example, the pleasure “Little Peter” derived from eating candy outweighed his fear of the rabbit because the rabbit was at first placed far enough away. If the rabbit had initially been placed too close to Peter, then he would presumably stop eating

and instead display signs of fear. In other words, Peter's aversive system would inhibit his appetitive system, as fear is incompatible with feeding. Indeed, this form of *conditioned suppression*, where appetitive responses are inhibited by the presence of a fear cue, is a common measure in conditioning preparations (Estes & Skinner, 1941). The idea that therapy should occur while the subject is in a non-anxious state is a key distinction between clinical treatments derived from counterconditioning (i.e., desensitization) and those derived from standard extinction (i.e., flooding or prolonged exposure therapy; Asnaani, McLean, & Foa, 2016). Indeed, relaxation is incompatible with some forms of exposure therapy that emphasize the importance of maximizing inhibitory or corrective learning (i.e. safety based meaning of a feared stimulus) during a state of anxiety (Abramowitz, 2013).

Counterconditioning versus standard extinction

Extinction refers both to the technique of omitting a US or reinforcement following conditioning, as well as to the process by which omission leads to a reduction in learned behavior. It is considered a form of retroactive inhibition that interferes with expression of the originally learned response. It is also widely appreciated that extinction tends to be a fairly weak and impermanent form of inhibition, and that the originally learned response often reemerges under a variety of circumstances (Mark E Bouton, 2002). Given the transient nature of extinction, there has long been motivation to strengthen retroactive inhibition using techniques theoretically derived from, but stronger than, standard forms of experimental extinction (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; Dunsmoor, Niv, Daw, & Phelps, 2015).

Results from Pavlov's laboratory (described in Joseph Wolpe & Plaud, 1997), as well as the "Little Peter" experiment, gave early promise to the effectiveness of counterconditioning techniques to reduce maladaptive fear. But in terms of persistently reducing the originally learned behavior, it has been questionable whether a counterconditioning approach is any more effective than simply omitting the US. Evidence that counterconditioning is more effective than standard extinction has been somewhat inconsistent in the early animal learning literature (Capaldi, Viveiros, & Campbell, 1983; Dickinson & Pearce, 1977; Richardson, Riccio, & Smoller, 1987). That is, some early studies failed to find any effects of counterconditioning on transforming behavior (e.g., Delprato & Jackson, 1973), and research that did show effects of counterconditioning was criticized for some critical methodological confounds (Lomont, 1965; Richardson et al., 1987; G. T. Wilson & Davison, 1971). A predominant methodological criticism of early counterconditioning research was that the role of standard extinction processes (distinct from counterconditioning) was often not sufficiently controlled for (Lomont, 1965; G. T. Wilson & Davison, 1971). Several early studies did not include an extinction-only group, nor did they control for the amount of exposure to the CS during counterconditioning. This is especially important in aversive-to-appetitive conditioning, because appetitive stimuli might simply encourage the subject to engage with the feared CS (or remain in a feared compartment) for a longer period of time while feeding or awaiting the arrival of food. Consequently, more time engaging with the CS provides more time for basic extinction processes to operate (G. T. Wilson & Davison, 1971). Counterconditioning could also have confounding influences on how much attention is paid to the CS during the second round of learning (Riccio, Richardson, & Ebner, 1984;

G. T. Wilson & Davison, 1971). For instance, the unexpected appearance of food (US2) in a previously shocked (US1) environment could reorient attention to the CS. Renewed attention to the CS could then provide a better opportunity for extinction mechanism to engage, or could accelerate new learning through an associability mechanism as proposed by influential attentional-associative learning models (Courville, Daw, & Touretzky, 2006; J. M. Pearce & Hall, 1980).

Below, we review non-human animal and human behavioral and neurobiological studies using counterconditioning techniques. (See Table 1 for a list of human counterconditioning studies). Much of this work overcomes earlier critiques on the field by including an extinction condition or a proper control group. Where appropriate, we highlight whether results show counterconditioning to be more effective, equally effective, or less effective than standard extinction.

Non-human animal research on counterconditioning

While the focus of counterconditioning has predominately been on modulating responses to the CS, early work showed that counterconditioning alters the value of a US (US1) when it is paired with another US of the opposite valence (US2). Pearce & Dickinson (1975) showed that pairing shock with food (US1-US2 pairing) reduced the effectiveness of the shock in subsequent conditioned suppression training. Another fairly robust finding in the counterconditioning literature is that of proactive interference. That is, animals are slower to acquire a new behavioral response if the CS had previously been associated with a US of the opposite valence (M. E. Bouton & Peck, 1992; J Konorski & Szwejkowska, 1956; Peck & Bouton, 1990; Scavio & Gormezano, 1980). In other words, counterconditioning requires more time to learn than de novo conditioning. This effect is similar to latent inhibition (Lubow, 1973), wherein repeated presentations of the CS without the US delays its ability to later form an association with the US.

A number of animal studies found that counterconditioning was superior to standard extinction. Early research showed that replacing an aversive US with food (Richardson, Riccio, Jamis, Cabosky, & Skoczen, 1982; Richardson et al., 1987; E. H. Wilson & Dinsmoor, 1970) or intra cranial self-stimulation of the hypothalamus (Reid, 1973) during exposure to fearful stimuli reduced fear-related behavior towards the CS more effectively than mere exposure. More recently, Thomas et al. (2012) explored the effectiveness of a modified counterconditioning procedure to reduce fear following fear conditioning. Unlike standard Pavlovian counterconditioning, where the reward is simply presented following the CS, rats had to perform an action (instrumental lever pressing) during CS presentation (lights off) to earn the reward (chocolate milk). Counterconditioning that involved earning the reward produced less contextual renewal than counterconditioning in which the reward (US2) was provided freely. Anderson et al. (2013) utilized novel objects to counter-condition learned fear in a passive avoidance task. Rats initially learned to avoid a compartment where they received mild shocks. The conditioned response was then extinguished by the omission of shock in the compartment, or counterconditioned by the presence of novel objects in the compartment (novel objects were considered naturally appetitive). During the testing session, in comparison to the standard extinction group, the counterconditioning group

exhibited less fear to the previously shocked compartment, as latency to cross to the feared compartment and time spent in the safe compartment were reduced. Correia et al. (2016) tested the effects of counterconditioning versus standard extinction on short term (immediate) and long-term (2 months) fear memory in rats. Rats underwent auditory fear conditioning followed by either standard extinction or counterconditioning, in which the CS (tone) was paired with sucrose delivery at a port chamber wall. Counterconditioning was more successful than standard extinction at reducing freezing behavior up to 2 months following training.

In contrast, a number of early studies showed counterconditioning to be equally as effective as extinction in reducing fear-related behavior (Capaldi et al., 1983; Delprato & Jackson, 1973; Klein, 1969; G. T. Wilson, 1973). Bouton and colleagues showed through a series of experiments that counterconditioning is consistent with other interference paradigms (extinction, latent inhibition, or reversal learning), such that the first association (CS-US1) and second association (CS-US2) later compete for behavioral expression. In this way, the effectiveness of counterconditioning is best assessed as performance at a future test when the CS is tested without paired US1 or US2. A series of experiments by Bouton and colleagues showed that counterconditioning slowed conditioning with the new US2, but that the original conditioned response established by CS-US1 pairings ultimately returned. Factors promoting the return of the originally learned behavior after counterconditioning includes when the CS is presented back in the original conditioning context (renewal; Peck & Bouton, 1990), as time elapse after counterconditioning (spontaneous recovery; M. E. Bouton & Peck, 1992), or when the original US1 is re-presented, unpaired, after counterconditioning (reinstatement; Brooks, Hale, Nelson, & Bouton, 1995). Importantly, the original behavior is prone to return in both aversive-to-appetitive or appetitive-to-aversive counterconditioning preparations (M. E. Bouton & Peck, 1992). Hence, the critical factor in guiding long-term behavior appears to be whichever behavior was learned first. This further demonstrates the challenge in sustaining long-term behavior modification across mental health disorders from either positive or negative valence systems.

A recent study (Holmes et al., 2016) showed that aversive-to-appetitive counterconditioning is prone to even stronger renewal of the original fear response than an extinction-only procedure when testing takes place in either the conditioning context (ABA renewal) or in a novel context (ABC renewal). One potential reason that counterconditioning is especially prone to contextual renewal is that the presence of the new US2 increases discrimination between separate experiences; that is, the CS acts as either a signal of threat or a signal for reward, depending on the context. As a consequence, the CS-US2 association becomes hyperspecific to the context in which that US2 was experienced, and behavior fails to generalize outside that context. The substantial dissimilarity in experiences with the CS across different contexts could also result in “state-splitting,” in which the original meaning of the CS is preserved despite new incompatible experiences [see: Optimizing counterconditioning].

Neurobiology of Counterconditioning

Whether the neural mechanism by which counterconditioning achieves its effects are distinct from the well-delineated neural circuitry of extinction is unclear. The switch from appetitive-to-aversive conditioning has been linked to reduced levels of activity in midbrain periaqueductal grey, as compared to aversive conditioning without prior appetitive conditioning (Nasser & McNally, 2012). Appetitive-to-aversive conditioning is also linked to increased activity in regions associated with prediction error signaling, including the thalamus, insular cortex, lateral amygdala, and the nucleus accumbens (Nasser & McNally, 2012). There is also emerging evidence for distinct neural populations coding for appetitive and aversive CSs within the same brain region. For example, neurons in the lateral habenula show opposing responses to CSs paired with either aversive or appetitive US (Matsumoto & Hikosaka, 2009). Recent neurobiological research using activity dependent neural tagging in mice also reveals distinct neural populations in the hippocampus (Chen et al., 2019; Ramirez et al., 2015) and amygdala (Tye, Stuber, de Ridder, Bonci, & Janak, 2008) coding for reward and aversive experiences.

Complicating the distinction between the neural mechanisms of counterconditioning versus standard extinction, however, is that the absence of an expected aversive US might itself constitute a rewarding event mediated by a prediction error signal in the mesocorticolimbic dopaminergic system (Kalisch, Gerlicher, & Duvarci, 2019; McNally, Johansen, & Blair, 2011) (Figure 2). The role of dopamine in fear extinction is coming into view. Extinction memory is strengthened by increases in dopamine through optogenetic enhancement (Salinas-Hernández et al., 2018), deep brain stimulation (Rodriguez-Romaguera et al., 2012), dopamine agonists (Haaker et al., 2013), and pairing extinction trials with reward, i.e., counterconditioning (Correia et al., 2016). Likewise, blocking dopamine activity in the nucleus accumbens impairs extinction (Holtzman-Assif, Laurent, & Westbrook, 2010). Aversive-to-appetitive counterconditioning could therefore be construed as a form of “rewarded” extinction, as it may preferentially engage dopamine projecting circuits that are already active in extinction learning thereby facilitating basic extinction processes.

Human Research on counterconditioning

Human research focused on counterconditioning has been sparse, but the topic has maintained some interest in the area of pain, fear, and changing learned preferences. The use of counterconditioning as a strategy to reduce fear of pain shows mixed results. Meulders et al. (2015) first fear-conditioned a movement (CS) using a painful shock (US), and then paired the movement with either monetary reward (counterconditioning group) or simply omitted the US (extinction group). At test, both the extinction and counterconditioning groups exhibited reductions of self-reported pain-related fear, but there was no added benefit of counterconditioning over standard extinction. Nevertheless, there was some evidence that counterconditioning changed the valence ratings of the CS to be more positive from pre-to-post testing relative to standard extinction.

One notable limitation to using money as a rewarding outcome in counterconditioning preparations is that, unlike a painful US, money is a secondary reinforcer. On top of this,

money is often symbolic (e.g., a dollar sign), and participants might vary in their belief that they will actually receive the money at the end of the experiment. Indeed, Meulders et al. (2015) noted that some participants did not find the reward manipulation credible; that is, they did not expect to actually receive the money signaled on the CS trials at the end of the experiment. These methodological issues can make it difficult to directly compare the animal counterconditioning learning literature using primary rewards (e.g., food) to human research using secondary reinforcers such as money.

Claes et al. (2014) found that introducing a reward with a painful movement did not reduce pain-related fear associated with the movement. However, reward did induce subjects to make more painful movements and subjects were less hesitant when making a painful movement. Thus, although associating reward with a painful action might not affect self-reported pain-related fear, the presence of reward might reduce avoidance of those pain-related actions. That reward seeking inhibits avoidance is in line with the idea that opposing valence systems antagonize and reciprocally inhibit the other (Dickinson & Pearce, 1977; Joseph Wolpe, 1954).

There have been recent attempts to compare the effects of counterconditioning against standard extinction using typical human Pavlovian fear conditioning preparation. In a between-subjects behavioral experiment, Raes & Raedt (2012) first paired a CS (picture of a face) with an aversive US (100 decibel white noise). Participants then underwent either standard extinction, appetitive counterconditioning with a baby laugh US, or neutral counterconditioning with a simple tone as the new outcome. Neither forms of counterconditioning were more effective than standard extinction on self-reported CS valence, US expectancy, or CS fear ratings. Nevertheless, counterconditioning (using both positive and neutral stimuli) did reduce negative evaluative responses as measured by an implicit affective priming task.

An important question in the animal learning literature has involved whether counterconditioning is more effective than extinction at diminishing relapse to the originally learned response via recovery, renewal, or reinstatement (Mark E Bouton, 2002). Some human research is now beginning to address this question using multi-day experiments. In a between-subjects multi-day behavioral experiment, Kang et al. (2018) first paired a fear-relevant CS (picture of a spider) with an aversive US (shock). The next day, participants underwent either standard extinction or counterconditioning in which the CS was paired with a positive image of a cartoon. Threat expectancy and CS valence ratings were collected on the third day during spontaneous recovery and US reinstatement tests. There was evidence of diminished threat expectancy during recovery and reinstatement tests in the counterconditioning group, but negative valence ratings of the CS was not different between the groups.

A notable limitation to the behavioral experiments reviewed so far is the limited use or lack of reported physiological data. One recent exception tested counterconditioning versus standard extinction on recovery and reinstatement of skin conductance responses (SCR) and fear-potentiated startle responses (van Dis, Hagenaars, Bockting, & Engelhard, 2019). In a between-subjects multi-day behavioral experiment, participants first learned the association

between a CS (neutral face) and an aversive shock. The next day, participants underwent either standard extinction, counterconditioning in which the CS was paired with a positive 6-second film clip, or extinction with unpaired presentations of the positive film clip. In contrast to Kang et al. (2018), which used a similar protocol, counterconditioning did improve CS valence ratings. One notable methodological difference between these experiments is the nature of the appetitive outcome used in counterconditioning; a positive film clip (van Dis et al., 2019) may be a more salient stimulus able to affect valence ratings more effectively than a static comic image (Kang et al., 2018). Even so, tests of spontaneous recovery and reinstatement one week later did not reveal any differential effect of counterconditioning versus standard extinction on SCRs, fear-potentiated startle, or shock expectancy.

Another recent study found that pairing a CS with a positive picture diminished SCRs in a 24-hour test relative to a CS that underwent standard extinction (Keller & Dunsmoor, In Press). This study used a within-subjects category-conditioning paradigm (Dunsmoor, Martin, & LaBar, 2012) that allowed for a test of episodic (recognition) memory for each category exemplar encoded during fear conditioning or extinction/counterconditioning (Dunsmoor & Kroes, 2019). Subjects explicitly recognized an equal number of CS exemplars (pictures of animals or tools) that had been encoded during fear conditioning and associated with a shock US, but remembered more category exemplars associated with a positive picture from counterconditioning (picture of animals or tools) than exemplars paired to omission of shock during standard extinction (pictures of tools or animals, respectively). This finding suggests that counterconditioning might strengthen both implicit and explicit memory for safety, possibly providing stronger memory retrieval competition against the fear memory to help diminish fear relapse. This result is interesting in light of another recent finding that episodic memory for information encoded during fear extinction is relatively weaker than information encoded during fear conditioning (Dunsmoor et al., 2018).

Neuroimaging studies of counterconditioning in humans are sparse, and have produced mixed findings. In a within-subjects design, Schweckendiek et al. (2013) found no difference in amygdala or striatal activity for disgust pictures that were paired or unpaired with monetary reward. There was also no effect of monetary reward on reducing disgust or arousal ratings of the disgust pictures. Notably, the CSs used in Schweckendiek et al., were negatively valenced (disgust images) prior to counterconditioning, which therefore limits comparability to studies that involve an initial conditioning phase. An fMRI study of aversive-to-appetitive counterconditioning (shock US to monetary reward) also failed to find statistically meaningful effects on SCRs or amygdala activity between a rewarded versus neutral CS (Bulganin, Bach, & Wittmann, 2014). However, the design of this experiment departed substantially from a standard Pavlovian counterconditioning design, making it difficult to establish whether counterconditioning was any more or less effective than standard Pavlovian extinction.

Experiments using counterconditioning to change the valence of phobic stimuli have also produced mixed results. de Jong et al. (2000) investigated whether counterconditioning reduces disgust toward spiders, which is considered a core factor in spider phobia. Women with spider phobia were assigned to either an exposure-only group, or a group where

exposure was combined with tasty-food items and the participant's favorite music. Both treatment conditions were effective at reducing avoidance of spiders, and increased the valence of spiders from before to after treatment. However, counterconditioning was no more effective than exposure without reward. This finding was in contrast to an earlier report that exposure with pleasing music was a more effective treatment that mere exposure for animal phobia (Eifert, Craill, Carey, & O'Connor, 1988)—though the difference in methodologies across these two studies make them challenging to directly compare.

A distinction between standard extinction and counterconditioning that may be especially relevant to human associative learning is the effect on expectancy-learning versus evaluative-learning (Baeyens & De Houwer, 1995) (Figure 1). Expectancy-learning arises when a contingency is established between a CS and a US, and the CS then becomes a reliable predictor for the appearance of the US. In other words, expectancy learning requires a contingency between the CS and US, and the absence of the US therefore diminishes expectancy. By contrast, evaluative-learning arises when the association between the CS and US changes the valence of the CS or preferences regarding the CS. Evaluative-learning can occur by mere contiguity between a CS and US and therefore the appearance of the CS does not have to indicate the presence of the US. As a consequence, evaluative conditioning can more easily survive extinction via the absence of the US.

Both expectancy and evaluative learning can co-occur in human conditioning preparations (Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002). However, behaviors associated with evaluative conditioning (e.g., changes in valence or preferences) are less sensitive to extinction than behaviors associated with expectancy-learning (e.g., defensive responses or fear ratings) (Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Baeyens, Eelen, Crombez, & Van den Bergh, 1992). These differences are perhaps due to the nature of extinction learning as predominately targeting the CS-US contingency. Because US omission does not affect behaviors learned by mere contiguity, the omission of the US might simply be insufficient to extinguish evaluative conditioning. In other words, evaluative conditioning is insensitive to expectations whether or not the US will appear (Baeyens, Eelen, & Crombez, 1995). Notably, evaluative conditioning appears more sensitive to the effects of counterconditioning than extinction in regards to changing CS valence and preferences (Baeyens, Eelen, Van den Bergh, & Crombez, 1989; Kerkhof et al., 2011). This effect could be due to verbal reports of CS valence reflecting integration of stimulus contingencies across the entire experiment (Lipp & Purkis, 2006). In this way, valence ratings after extinction would be influenced by a history of reinforced and unreinforced trials, whereas valence ratings after counterconditioning would integrate a history of appetitive and aversive outcomes.

In the tradition of the "Little Peter" experiment (Jones, 1924), an active area of counterconditioning research continues to be minimizing childhood fears. Counterconditioning in children has shown to be more effective than standard extinction at reducing fear beliefs and avoidance (Newall, Watson, Grant, & Richardson, 2017) and decrease heart-rate responding (Reynolds, Field, & Askew, 2018). Providing children with positive information about a feared CS has also been shown to reduce fear more effectively

than modelling a non-anxious response towards the fear CS (Kelly, Barker, Field, Wilson, & Reynolds, 2010; Muris, Huijding, Mayer, van As, & van Alem, 2011).

In summary, there is evidence that counterconditioning is more effective at reducing fear than standard extinction in laboratory animals and in humans. A positive stimulus during extinction may serve to change the valence of the CS (Engelhard, Leer, Lange, & Olatunji, 2014; Kerkhof et al., 2011; Raes & De Raedt, 2012), but there is to date only limited evidence that aversive-to-appetitive counterconditioning persistently changes physiological measures of fear in human adults. Overall, the noticeable limit of basic research in counterconditioning relative to standard extinction is noteworthy, given the possibility that counterconditioning may be more effective and can be utilized in a therapeutic context to improve treatment outcomes for a variety of psychiatric disorders.

Derivatives of counterconditioning

Another approach to augment extinction training involves replacing a valenced US (e.g., shock) with a neutral stimulus. Raes & Raedt (2012) found that replacing the aversive US with a neutral outcome (a simple tone) was equivalent to replacing the aversive US with an appetitive outcome (the sound of baby laughter), in reducing evaluative responses of fear. In a cross-species behavioral experiment, Dunsmoor et al. (2015) tested the effect of augmenting extinction by replacing shocks with tones (what they referred to as *novelty-facilitated extinction*), and compared that technique against standard extinction that involved merely omitting the shocks. In both humans and in rats, replacing shocks with a simple tone during extinction resulted in less spontaneous recovery of SCRs in humans and freezing in rats when testing occurred in the absence of shocks or tones. These behavioral results were replicated in humans in an fMRI experiment (Dunsmoor et al., 2019), and novelty-facilitated extinction was recently found to diminish reinstatement (Lucas, Luck, & Lipp, 2018). Krypotos & Engelhard (2018) did not find that replacing shocks with neutral outcomes was any more effective than standard extinction on reinstatement of avoidance behavior or explicit ratings of self-reported fear. However, an avoidance design differs in a number of critical aspects from Pavlovian conditioning (LeDoux, Moscarello, Sears, & Campese, 2017), which can complicate drawing parallels in how these behaviors are extinguished.

An important feature shared by both counterconditioning and novelty-facilitated extinction is the presence of a new perceptible outcome, as opposed to simply the omission of the expected outcome as in extinction. Like conditioning, extinction has long been considered a form of new learning (Pavlov, 1927; J. M. Pearce & Hall, 1980) that requires a prediction error between the expected and received outcome. Rescorla & Wagner (1972) initially described the negative prediction error as effectively erasing associative value of the CS. But given that extinction is impermanent, most associative learning models view extinction as generating a new association between the CS and the absence of the US (e.g., Larrauri & Schmajuk, 2008; J. M. Pearce & Hall, 1980) (sometimes referred to as the “CS-no US” association), with a number of factors determining which CS association is later retrieved (M. E. Bouton, 1993). The presence of a new outcome (either a US2 or a neutral outcome like a tone) might generate a stronger prediction error to putatively drive new associative learning than the mere omission of shock. In experimental preparations, this might occur

when the US is omitted on some CS trials during learning. Indeed, conditioning that involves fewer USs than CSs (partial reinforcement) is paradoxically harder to extinguish than conditioning with an equal number of CSs and USs (continuous reinforcement), known as the partial reinforcement extinction effect (Humphreys, 1939). Likewise, in real world scenarios the absence of a feared outcome is oftentimes insufficient to disconfirm threat expectancy, nor do encounters with a phobic stimulus itself constitute effective exposure treatment (Craske et al., 2014). For example, people with a fear of public speaking may continue to dread giving public presentations despite repeated speaking engagements.

Because new learning relies on detecting the discrepancy between the expected and received outcome, extinction learning might therefore suffer in cases when the absence of the US is insufficient to drive new learning. Thus, a mechanism underlying the effectiveness of counterconditioning (and novelty-facilitated extinction) may be in producing a stronger prediction error than standard extinction to signal that the CS-US contingencies have changed and are in need of updating. In this framework, providing a new outcome (as opposed to merely omitting the old outcome) might accelerate new learning in accordance with computational models of associability (Courville et al., 2006; J. M. Pearce & Hall, 1980; Roesch, Esber, Li, Daw, & Schoenbaum, 2012). Associability is a feature of a CS akin to attentional gating that dynamically determines (i.e., trial-to-trial) how susceptible the CS is to new associative learning. Associability increases with surprise, such as the first appearance of the US, and diminishes as the outcome becomes less surprising, such as over the course of CS-US training. This same mechanism occurs during extinction, putatively increasing associability when the US is surprisingly omitted, thereby driving new CS-no US learning. Providing a new outcome might simply generate a stronger prediction error to restore associability to the CS, opening a window of opportunity to accelerate new learning. Indeed, Dunsmoor et al. (2019) found that SCRs and ventromedial prefrontal cortex activity during novelty-facilitated extinction conformed to a version of a Pearce-Hall associability model. Moreover, subjects who showed faster within-session updating of associability showed less fear recovery the next day.

One important conceptual distinction between aversive-to-appetitive counterconditioning and novelty-facilitated extinction might concern maintaining the motivational significance of the CS. That is, in counterconditioning the motivational significance of the CS is maintained, as it remains a signal for a salient outcome. Indeed, counterconditioning is sometimes referred to in the literature as “cross-motivational transfer” (M. E. Bouton & Peck, 1992). Because the CS is associated with two competing behavioral responses, the subject is faced with retrieving a US representation to guide the appropriate behavior, such as whether to approach or avoid the CS. By contrast, replacing the shock with a neutral outcome does not command a new competing response or motivationally pertinent US2 representation—the subject could learn to ignore the CS as trivial. Thus the presence of the novel outcome might maintain attention to the CS, at least in the short term, thereby increasing associability, while US1 omission allows extinction processes to fully engage. Moreover, subjects might habituate more rapidly both to the CS and the outcome in novelty-facilitated extinction than in counterconditioning, given that the outcome is neutral, which may lead to different effects from these two approaches. Further research is needed to

evaluate the effects of novelty-facilitated extinction as compared to counterconditioning and extinction, both empirically and from a theoretical and conceptual standpoint.

Another technique related to counterconditioning involves positive valence training. Dour et al. (2016) exposed individuals with a fear of spiders to a tarantula, followed by a 7 minute film clip that either described the positive aspects of spiders or a control video unrelated to spiders. The positive valence training group showed less behavioral avoidance and less negative valence toward spiders that persisted for up to 2 weeks. In a human fear conditioning experiment, Luck & Lipp (2018) showed that positive information about a fear-conditioned CS (image of a person) and negative information about a control CS reversed CS valence ratings. But positive valence training did not affect reinstatement of SCRs. A related form of positive valence training involves conducting extinction training immediately after inducing a positive mood. For example, Zbozinek et al. (2015) conducted a human fear conditioning experiment using neutral CS (face) paired with a shock US followed by extinction. Immediately after fear conditioning and just prior to extinction, participants underwent a positive mood induction or a control task. Positive mood induction increased CS valence ratings following extinction, relative to a control condition, and reduced reinstatement of fear-potentiated startle and self-reported fear at a one week test.

The clinical relevance of counterconditioning

Negative valence system disorders

The principles of counterconditioning have been effectively utilized through systematic desensitization therapy to treat disorders characterized by negative valence, such as Specific Phobias and Obsessive Compulsive Disorder (Tryon, 2005). Paunovic (2011) developed a prolonged exposure counterconditioning method to treat PTSD by asking patients to undergo imaginal reliving of a pleasurable emotional experience incompatible to the trauma. Approaches that integrate exposure treatment with appetitive counterconditioning may be useful to prevent dropout or encourage treatment for patients who find prolonged exposure therapy difficult to tolerate. At a broader level, any effective treatment might presumably contain elements of counterconditioning as long as patients derive satisfaction from overcoming their fear. The underlying mechanism of systematic desensitization therapy, as originally proposed by Wolpe, emphasized the importance of relaxation during gradual imaginal presentations of feared stimuli to inhibit anxiety during treatment. However, whether relaxation is a key element to the long-term success of extinction-based treatments for fear reduction remains contested (Abramowitz, 2013; Borkovec & Sides, 1979).

Proactive interference, like that seen in counterconditioning (M. E. Bouton & Peck, 1992; Dickinson & Pearce, 1977) or latent inhibition (Lubow, 1973), has also been harnessed in a clinical setting as a form of resilience training to prepare patients for future stressful events. In Stress Inoculation Training (Maag & Kotlash, 1994; Meichenbaum, 1985), patients are prepared (“inoculated”) for the potential negative psychological impact of an imminent stressful life event through a variety of approaches. One approach involves exposing clients to milder forms of stress in order to bolster both coping mechanisms and the individual’s confidence in using his or her coping repertoire (Serino et al., 2014). This training has also been used as a preventative strategy in combination with virtual reality to increase coping

strategies for military personnel, students in an earthquake scenario and employees in a hostile work environment (Wiederhold & Wiederhold, 2008). The combination of mental health therapy and virtual reality has shown promise across a number of domains (Powers & Emmelkamp, 2008), and may be useful for generating enriched positive environments for counterconditioning derived treatment.

One unusual problem with aversive-to-appetitive counterconditioning is if the appetitive US is excessively rewarding. That is, as the reward is typically omitted at test (and also likely in the real-world), the absence of reward could lead to disappointment or even frustration (Burokas, Gutiérrez-Cuesta, Martín-García, & Maldonado, 2012). Indeed, frustration may impair the results of counterconditioning (Capaldi et al., 1983; Dickinson & Pearce, 1977). For clinical purposes, it may be important to consider the nature of reward used in counterconditioning treatment, such that absence of reward in the real-world does not lead to disappointment and frustration, which in turn can lead to relapse of the original maladaptive behavior.

Positive valence system disorders

Contemporary laboratory research on appetitive-to-aversive counterconditioning in humans, and its translational value, is especially rare. In an aversive counterconditioning study in humans, Van Gucht et al. (2010) initially paired a CS (serving tray) with chocolate, and then extinguished the appetitive association by presenting the CS without chocolate in one group or by pairing it with the consumption of a highly disliked liquid. In comparison to standard extinction, aversive conditioning was able to more effectively reduce expectancy and liking of the appetitive US. In an fMRI study of appetitive-to-aversive counterconditioning, Kaag et al. (2016) first paired two CSs (colored squares) with monetary reward, and then extinguished one CS and paired the other CS with a shock US. Following reinstatement of the monetary reward US, activity in the ventral striatum and ventral tegmental area was reduced to the aversively counter-conditioned CS as compared to the extinguished CS. The results indicate a potential neural correlate for aversion therapy, whereby the rewarding value of a CS is altered via a secondary association between the CS and a negative outcome.

The clinical analogue to appetitive-to-aversive counterconditioning is aversion therapy, in which the goal is to re-associate maladaptive reward-seeking behaviors with negative outcomes or punishment. Aversion therapy has a controversial history in clinical practice. It is perhaps most popularly recognized from a disturbing scene in Stanley Kubrick's *A Clockwork Orange*, and it has been used in conversion therapy to attempt to change sexual orientation. A more common use for aversion-based therapy involves medications that render addictive substances less pleasurable, or render them extremely displeasing. For example, the pharmacological agent disulfiram has been used to treat alcohol use disorder by generating a highly unpleasant physical and psychological reaction when combined with alcohol (Fuller et al., 1986). But the effectiveness of disulfiram in preventing relapse in unsupervised settings is unestablished (M. D. Skinner, Lahmek, Pham, & Aubin, 2014). Naltrexone, a competitive opioid receptor antagonist, is likewise used to treat alcohol use disorder as well as opioid addiction by reducing the reinforcing properties of the drugs. Naltrexone can be part of an effective treatment, but issues of adherence to a treatment plan

that eliminates the appetitive properties of an addictive drug outside a supervised setting, similarly arise (Carmen, Angeles, Ana, & María, 2004).

Optimizing counterconditioning

The ultimate goal of seeking alternative approaches to standard extinction is to innovate clinical treatment to more effectively prevent the relapse of unwanted thoughts and behavior. Counterconditioning is procedurally distinct from standard extinction because it involves the presentation of a new outcome, rather than omission of the expected outcome. Mechanistically, counterconditioning might operate by increasing the prediction error, thereby enhancing the recruitment of dopaminergic motivational systems, which in turn augments the function of standard extinction processes. Yet, as discussed, basic rodent and human research on counterconditioning is extremely mixed, with many studies showing that the originally learned CS-US1 association is prone to relapse. Next, we discuss strategies to optimize counterconditioning to prevent relapse.

Holmes et al. (2016) found that counterconditioning was more prone to fear renewal than extinction, possibly because the presence of US2 during counterconditioning increased discrimination between contexts. Contextual renewal is an innate problem with extinction, or any manipulation where the CS is not presented in conjunction with its original US1. That is, the second experience can be interpreted as an exception to an important rule, and the lack of the expected US1 can generate a degree of uncertainty or ambiguity. Some strategies have been adopted to enhance the generalization of extinction by reducing the effects of context shifts on return of the original CR. Gershman et al. (2013) tried to prevent the abrupt shift between competing states of conditioning and extinction using a *gradual extinction* technique, where extinction trials were reinforced with a US at a diminishing rate. They showed that rodents that underwent gradual extinction exhibited less spontaneous recovery and fear reinstatement than those that underwent standard extinction. A similar technique may prove effective for counterconditioning, perhaps by gradually interweaving US2 trials with US1 trials during extinction (see also Woods & Bouton, 2007).

Conducting extinction in multiple contexts is another way to possibly prevent fear relapse. In comparison to standard extinction, studies have shown that multiple context extinction can reduce renewal (Gunther, Denniston, & Miller, 1998; Shiban, Pauli, & Muhlberger, 2013) and reinstatement (Dunsmoor, Ahs, Zielinski, & LaBar, 2014). By conducting secondary learning (e.g., extinction or counterconditioning) in several different contexts, new learning may generalize more easily to additional contexts in the future. Repeatedly changing context exposures might also enhance novelty during new learning that, together with the presence of a reward, enhances learning and attention during a counterconditioning phase.

Another approach to optimize the effects of counterconditioning is to incorporate an unexpected reward into a memory reconsolidation updating framework. Reconsolidation refers to the theory that reactivation of a long-term memory brings it to a labile state where it is sensitive to change (Alberini & LeDoux, 2013). The past two decades has seen excitement and disappointment in whether neurobiological models of memory reconsolidation can be translated to clinical treatment for mental health disorders governed by fear and anxiety.

Despite early promise, utilizing a reconsolidation framework with pharmacological treatment, such as beta-blockers, to alleviate the symptoms of PTSD has shown mixed and limited efficacy (Brunet et al., 2008; Giustino, Fitzgerald, & Maren, 2016; Sharp, Thomas, Rosenberg, Rosenberg, & Meyer III, 2010; Wood et al., 2015). A non-pharmacological strategy has been developed that involves reactivating a fear memory via an isolated reminder trial prior to an extinction session (retrieval + extinction) (Monfils, Cowansage, Klann, & LeDoux, 2009). Interestingly, this general technique was pre-dated by Richardson et al. (1982), where they found that presenting a maltose solution following a contextual fear memory reactivation session decreased fear more effectively than memory reactivation without reward, or no memory reactivation. More recently, Haubrich et al. (2015) showed that providing a positive US (chocolate) while a fear memory was reactivated diminished the return of fear. Interestingly, there was no effect if the fear memory was reactivated in the presence of a non-edible neutral stimulus. Pedraza et al. (2018) showed that caffeine administration induces anxiolytic responses in rodents and can reduce fear relapse if presented with a CS during fear memory reactivation.

The reconsolidation period of a memory has also been harnessed to prevent the relapse of drug seeking behavior. Goltseker et al. (2017) tested the effects of aversive counterconditioning during the reconsolidation period of a contextual cocaine memory. To compare the effects of counterconditioning on memory reconsolidation, one group underwent memory retrieval before aversive counterconditioning (administration of lithium chloride) while the other group did not. In the memory retrieval group, mice were confined to the paired compartment for a short time and then placed in their home cage. Afterwards, mice were administered lithium chloride before being confined to the paired compartment. In the no retrieval group, mice were not exposed to the paired compartment before lithium chloride administration. In comparison to the no retrieval group, counterconditioning after memory retrieval was able to reduce reinstatement of conditioned place preference to the drug paired compartment, 24 hours after reconsolidation. Retrieval of an aversive lithium chloride contextual memory followed by appetitive counterconditioning with cocaine, prevented reinstatement of lithuim chloride place aversion. Thus counterconditioning following memory retrieval can alter both appetitive and aversive memories.

The effect of counterconditioning on memory reconsolidation updating is an exciting avenue of research that warrants further investigation. The few studies in this area are in line with the idea that an adaptive purpose of reconsolidation is to incorporate new information at the time of retrieval (Lee, 2009). Thus, stimuli of the opposite valence might be utilized to retune and re-encode an aversive or appetitive memory. But failures to reduce learned behavior using a reconsolidation framework can be due to several boundary conditions, such as the strength and age of the memory, and procedural differences established by different protocol designs (Treanor, Brown, Rissman, & Craske, 2017). For example, a 30 day old fear memory (Haubrich et al., 2015) and a strong fear conditioning memory (Pedraza et al., 2018) were not affected by a standard reconsolidation + counterconditioning procedure. And there have been notable failures to replicate the retrieval + extinction effect in rats and humans that limit the potential clinical utility of this strategy for treating psychiatric disorders (Chan, Leung, Westbrook, & McNally, 2010; Golkar, Bellander, Olsson, & Ohman, 2012; Kroes, Dunsmoor, Lin, Evans, & Phelps, 2017; Luyten & Beckers, 2017)

Counterconditioning via real-time fMRI neurofeedback

An inherit drawback of exposure-based treatments is the distress and anxiety caused by stimulus exposure during treatment. In the worst case, the anxiety produced during exposure leads to patient attrition (Loerinc et al., 2015). Real-time fMRI neurofeedback (rt-fMRI) combined with multivariate pattern analysis (MVPA) may offer one path of treatment without the need for patient exposure to the stimulus (Norman, Polyn, Detre, & Haxby, 2006; Sitaram et al., 2017). The general goal of combining these neuroimaging techniques is to encourage participants to activate neural representations of feared stimuli without US reinforcement, in effect achieving stimulus exposure without a perceptible stimulus. Koizumi and et al. (2016) first explored this method by conditioning subjects to two different CS+ stimuli consisting of colored vertical gratings, and then randomly choosing one CS+ stimulus as the target for subsequent rt-fMRI neurofeedback sessions. Participants completed a neurofeedback task in which they were instructed to increase the diameter of a circle, the size of which corresponded to a monetary reward, using any possible method. Neurofeedback scores were calculated as MVPA classifier probability of the evoked target CS+ pattern in visual cortex. When CS+ fear responses were tested after 3 neurofeedback sessions, participants had significantly lower SCR and amygdala responses to the neurofeedback targeted CS+ compared to the non-target CS+.

Although this technique is a promising cutting-edge approach to exposure therapy, implementation of this specific method still requires participants to initially see the feared CS+ in order to identify patterns of neural activity associated with perceiving it in the first place. Another approach from Taschereau-Dumouchel et al. (2018) expands on this idea of rt-fMRI neurofeedback in two important ways: by using participants with sub-clinical and clinical phobias as opposed to simple conditioning, and by using hyperalignment to construct a between subjects MVPA classifier. Hyperalignment refers to an MVPA strategy in which a classifier can be trained on multiple subjects and used to predict novel neural representations in a different subject (Haxby, 2011). Specifically, experimenters trained a hyperaligned MVPA classifier on representations of feared stimuli in non-phobic subjects, and were then able to target these representations for neurofeedback in phobic subjects. This approach is ideal for treatment of fear, as it removes the necessity for patients having to view a feared stimulus in order to construct an MVPA classifier. Using a standard visual feedback and monetary reward rt-fMRI paradigm, a sub-clinical and a small clinical cohort of subjects were able to reduce SCR and amygdala activity without having to undergo direct exposure to images of a phobic animal stimulus.

In these studies, participants remained unaware as to both the target stimulus and goal of the neurofeedback training. Subjects remained consciously unaware of the pairing of reward with the neural representation corresponding to the feared stimulus, thus constituting a unique form of counterconditioning to reduce behavioral and neural threat responses. An information transmission analysis was used in each study to investigate which brain regions outside of visual cortex tracked the probability of inducing the target CS pattern during neurofeedback. Both studies found significant disengagement of the vmPFC during neurofeedback, while Koizumi et al. (2016) also showed engagement of the striatum. In each case, the authors suggest that this paradoxical downregulation of the vmPFC during

neurofeedback counterconditioning implicates a different neural mechanisms of fear reduction from one involving canonical extinction neurocircuitry, and possibly involving neural reward circuits. Real-time fMRI feedback thus provides a promising avenue of fear reduction via counterconditioning without stimulus presentation, however the neural mechanisms underlying these processes remain unclear.

Conclusion

Historically, research on counterconditioning has been largely situated in learning theory accounts of animal behavior. But research on affective value coding in the brain has borrowed from the theoretical tradition of aversive-appetitive interactions in Pavlovian conditioning, and exciting avenues of neuroimaging research in humans is using counterconditioning in an attempt to change the neural representations of feared stimuli. As research on counterconditioning moves forward, an important question remains: are counterconditioning and extinction distinct processes, or is counterconditioning a means to facilitate extinction? Further research and development in this area will be useful to optimize clinical treatments and prevent relapse of maladaptive behavior.

ACKNOWLEDGMENTS

The authors acknowledge funding from NIH R00MH106719 and a National Science Foundation CAREER award (#1844792) to JED and NIH R00MH106719-04S to NEK.

References

Abramowitz JS (2013). The practice of exposure therapy: relevance of cognitive-behavioral theory and extinction theory. Behavior Therapy, 44(4), 548–558. [PubMed: 24094780]

Alberini CM, & LeDoux JE (2013). Memory reconsolidation. Current Biology, 23(17), R746–750. doi: 10.1016/j.cub.2013.06.046 [PubMed: 24028957]

Anderson MJ, Burpee TE, Wall MJ, & McGraw JJ (2013). Exposure to Novelty Weakens Conditioned Fear in Long-Evans Rats. Psicologica: International Journal of Methodology and Experimental Psychology, 34(1), 59–78.

Asnaani A, McLean CP, & Foa EB (2016). Updating Watson & Marks (1971): How has our understanding of the mechanisms of extinction learning evolved and where is our field going next? Behavior Therapy, 47(5), 654–668. [PubMed: 27816078]

Baeyens F, Crombez G, Van den Bergh O, & Eelen P (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. Advances in behaviour research and therapy, 10(4), 179–199.

Baeyens F, & De Houwer J (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994). Behaviour Research and Therapy, 33(7), 825–831. [PubMed: 7677721]

Baeyens F, Eelen P, & Crombez G (1995). Pavlovian associations are forever: On classical conditioning and extinction. Journal of Psychophysiology.

Baeyens F, Eelen P, Crombez G, & Van den Bergh O (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. Behaviour Research and Therapy, 30(2), 133–142. [PubMed: 1567342]

Baeyens F, Eelen P, Van den Bergh O, & Crombez G (1989). Acquired affective-evaluative value: Conservative but not unchangeable. Behaviour Research and Therapy, 27(3), 279–287. [PubMed: 2730509]

Berridge KC (2019). Affective valence in the brain: modules or modes? Nature Reviews Neuroscience.

Borkovec TD, & Sides JK (1979). The contribution of relaxation and expectancy to fear reduction via graded, imaginal exposure to feared stimuli. Behaviour Research and Therapy, 17(6), 529–540. [PubMed: 43126]

Bouton ME (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. Psychological bulletin, 114(1), 80–99. [PubMed: 8346330]

Bouton ME (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. Biological psychiatry, 52(10), 976–986. [PubMed: 12437938]

Bouton ME (2004). Context and behavioral processes in extinction. Learning & Memory, 11(5), 485494. doi:10.1101/lm.78804

Bouton ME (2014). Why behavior change is difficult to sustain. Preventive Medicine, 68, 29–36. doi: 10.1016/j.ypmed.2014.06.010 [PubMed: 24937649]

Bouton ME, & Peck CA (1992). Spontaneous recovery in cross-motivational transfer (counterconditioning). Animal Learning & Behavior, 20(4), 313–321. doi:10.3758/bf03197954

Brooks DC, Hale B, Nelson JB, & Bouton ME (1995). Reinstatement after counterconditioning. Animal Learning & Behavior, 23(4), 383–390. doi:10.3758/bf03198938

Brunet A, Orr SP, Tremblay J, Robertson K, Nader K, & Pitman RK (2008). Effect of post-retrieval propranolol on psychophysiologic responding during subsequent script-driven traumatic imagery in post-traumatic stress disorder. Journal of psychiatric research, 42(6), 503–506. doi:10.1016/j.jpsychires.2007.05.006 [PubMed: 17588604]

Bulganin L, Bach DR, & Wittmann BC (2014). Prior fear conditioning and reward learning interact in fear and reward networks. Frontiers in Behavioral Neuroscience, 8, 67. [PubMed: 24624068]

Burokas A, Gutiérrez-Cuesta J, Martín-García E, & Maldonado R (2012). Operant model of frustrated expected reward in mice. Addiction biology, 17(4), 770–782. [PubMed: 22264360]

Capaldi ED, Viveiros DM, & Campbell DH (1983). Food as a contextual cue in counterconditioning experiments: Is there a counterconditioning process? Animal Learning & Behavior, 11(2), 213–222.

Carmen B, Angeles M, Ana M, & María AJ (2004). Efficacy and safety of naltrexone and acamprosate in the treatment of alcohol dependence: a systematic review. Addiction, 99(7), 811–828. [PubMed: 15200577]

Chan WYM, Leung HT, Westbrook RF, & McNally GP (2010). Effects of recent exposure to a conditioned stimulus on extinction of Pavlovian fear conditioning. Learning & Memory, 17(10), 512–521. doi:10.1101/lm.1912510 [PubMed: 20884753]

Chen BK, Murawski NJ, Cincotta C, McKissick O, Finkelstein A, Hamidi AB, … Shpokayte M (2019). Artificially Enhancing and Suppressing Hippocampus-Mediated Memories. Current Biology, 29(11), 1885–1894. e1884. [PubMed: 31130452]

Claes N, Karos K, Meulders A, Crombez G, & Vlaeyen JW (2014). Competing goals attenuate avoidance behavior in the context of pain. The Journal of Pain, 15(11), 1120–1129. [PubMed: 25152301]

Correia SS, McGrath AG, Lee A, Graybiel AM, & Goosens KA (2016). Amygdala-ventral striatum circuit activation decreases long-term fear. eLife, 5, e12669. [PubMed: 27671733]

Courville AC, Daw ND, & Touretzky DS (2006). Bayesian theories of conditioning in a changing world. Trends in cognitive sciences, 10(7), 294–300. doi:10.1016/j.tics.2006.05.004 [PubMed: 16793323]

Craske MG, Treanor M, Conway CC, Zbozinek T, & Vervliet B (2014). Maximizing exposure therapy: An inhibitory learning approach. Behaviour Research and Therapy, 58C, 10–23. doi:10.1016/j.brat.2014.04.006

Daw ND, Kakade S, & Dayan P (2002). Opponent interactions between serotonin and dopamine. Neural Networks, 15(4), 603–616. [PubMed: 12371515]

de Jong PJ, Vorage I, & van den Hout MA (2000). Counterconditioning in the treatment of spider phobia: Effects on disgust, fear and valence. Behaviour Research and Therapy, 38(11), 1055–1069. [PubMed: 11060935]

Delprato DJ, & Jackson DE (1973). Counterconditioning and exposure-only in the treatment of specific (conditioned suppression) and generalized fear in rats. Behaviour Research and Therapy, 11(4), 453–461. [PubMed: 4798233]

Dickinson A, & Pearce JM (1977). Inhibitory interactions between appetitive and aversive stimuli. Psychological bulletin, 84(4), 690.

Dour HJ, Brown LA, & Craske MG (2016). Positive valence reduces susceptibility to return of fear and enhances approach behavior. Journal of behavior therapy and experimental psychiatry, 50, 277–282. [PubMed: 26497447]

Dunsmoor JE, Ahs F, Zielinski DJ, & LaBar KS (2014). Extinction in multiple virtual reality contexts diminishes fear reinstatement in humans. Neurobiology of learning and memory. doi:10.1016/j.nlm.2014.02.010

Dunsmoor JE, Campese VD, Ceceli AO, LeDoux JE, & Phelps EA (2015). Novelty-facilitated extinction: providing a novel outcome in place of an expected threat diminishes recovery of defensive responses. Biological psychiatry, 78(3), 203–209. [PubMed: 25636175]

Dunsmoor JE, & Kroes MC (2019). Episodic memory and Pavlovian conditioning: ships passing in the night. Current opinion in behavioral sciences, 26, 32–39. [PubMed: 31131296]

Dunsmoor JE, Kroes MC, Li J, Daw ND, Simpson HB, & Phelps EA (2019). Role of human ventromedial prefrontal cortex in learning and recall of enhanced extinction. Journal of Neuroscience, 39(17), 3264–3276. [PubMed: 30782974]

Dunsmoor JE, Kroes MC, Moscatelli CM, Evans MD, Davachi L, & Phelps EA (2018). Event segmentation protects emotional memories from competing experiences encoded close in time. Nature Human Behaviour, 2(4), 291.

Dunsmoor JE, Martin A, & LaBar KS (2012). Role of conceptual knowledge in learning and retention of conditioned fear. Biological Psychology, 89(2), 300–305. doi:10.1016/j.biopsycho.2011.11.002 [PubMed: 22118937]

Dunsmoor JE, Niv Y, Daw ND, & Phelps EA (2015). Rethinking extinction. Neuron, 88, 47–63. [PubMed: 26447572]

Eifert GH, Craill L, Carey E, & O'Connor C (1988). Affect modification through evaluative conditioning with music. Behaviour Research and Therapy, 26(4), 321–330. [PubMed: 3214396]

Engelhard IM, Leer A, Lange E, & Olatunji BO (2014). Shaking that icky feeling: Effects of extinction and counterconditioning on disgust-related evaluative learning. Behavior Therapy, 45(5), 708–719. [PubMed: 25022781]

Estes WK, & Skinner BF (1941). Some quantitative properties of anxiety. Journal of Experimental Psychology, 29(5), 390.

Fuller RK, Branchey L, Brightwell DR, Derman RM, Emrick CD, Iber FL, … Lowenstam I (1986). Disulfiram treatment of alcoholism: A Veterans Administration cooperative study. Jama, 256(11), 1449–1455. [PubMed: 3528541]

Gershman SJ, Jones CE, Norman KA, Monfils M-H, & Niv Y (2013). Gradual extinction prevents the return of fear: implications for the discovery of state. Frontiers in Behavioral Neuroscience, 7.

Giustino TF, Fitzgerald PJ, & Maren S (2016). Revisiting propranolol and PTSD: memory erasure or extinction enhancement? Neurobiology of learning and memory, 130, 26–33. [PubMed: 26808441]

Golkar A, Bellander M, Olsson A, & Ohman A (2012). Are fear memories erasable?-reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli. Frontiers in Behavioral Neuroscience, 6. doi:10.3389/fnbeh.2012.00080

Goltseker K, Bolotin L, & Barak S (2017). Counterconditioning during reconsolidation prevents relapse of cocaine memories. Neuropsychopharmacology, 42(3), 716. [PubMed: 27468918]

Grossberg S (2000). The imbalanced brain: From normal behavior to schizophrenia. Biological psychiatry, 48(2), 81–98. [PubMed: 10903405]

Gunther LM, Denniston JC, & Miller RR (1998). Conducting exposure treatment in multiple contexts can prevent relapse. Behaviour Research and Therapy, 36(1), 75–91. [PubMed: 9613018]

Haaker J, Gaburro S, Sah A, Gartmann N, Lonsdorf TB, Meier K, … Kalisch R (2013). Single dose of L-dopa makes extinction memories context-independent and prevents the return of fear. Proc Natl Acad Sci U S A, 110(26), E2428–2436. doi:10.1073/pnas.1303061110 [PubMed: 23754384]

Haubrich J, Crestani AP, Cassini LF, Santana F, Sierra RO, de O Alvares L, & Quillfeldt JA (2015). Reconsolidation allows fear memory to be updated to a less aversive level through the incorporation of appetitive information. Neuropsychopharmacology, 40(2), 315. [PubMed: 25027331]

Hermans D, Vansteenwegen D, Crombez G, Baeyens F, & Eelen P (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. Behaviour Research and Therapy, 40(3), 217–234. [PubMed: 11863234]

Holmes NM, Leung HT, & Westbrook RF (2016). Counterconditioned fear responses exhibit greater renewal than extinguished fear responses. Learning & Memory, 23(4), 141–150. [PubMed: 26980781]

Holtzman-Assif O, Laurent V, & Westbrook RF (2010). Blockade of dopamine activity in the nucleus accumbens impairs learning extinction of conditioned fear. Learning & Memory, 17(2), 71–75. [PubMed: 20154351]

Hull CL (1943). Principles of Behavior. New York: Appleton-Century-Crofts.

Humphreys LG (1939). The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. Journal of Experimental Psychology, 25(2), 141.

Jones MC (1924). A laboratory study of fear: The case of Peter. The Journal of Genetic Psychology, 31, 308–315.

Kaag AM, Schluter RS, Karel P, Homberg J, van den Brink W, Reneman L, & van Wingen GA (2016). Aversive counterconditioning attenuates reward signaling in the ventral striatum. Frontiers in human neuroscience, 10, 418. [PubMed: 27594829]

Kalisch R, Gerlicher AM, & Duvarci S (2019). A Dopaminergic Basis for Fear Extinction. Trends in cognitive sciences, 23(4), 274–277. [PubMed: 30803871]

Kang S, Vervliet B, Engelhard IM, van Dis EA, & Hagenaars MA (2018). Reduced return of threat expectancy after counterconditioning versus extinction. Behaviour Research and Therapy, 108, 78–84. [PubMed: 30064009]

Keller NE, & Dunsmoor JE (In Press). The effects of aversive-to-appetitive counterconditioning on implicit and explicit fear memory. Learning & Memory.

Kelly VL, Barker H, Field AP, Wilson C, & Reynolds S (2010). Can Rachman's indirect pathways be used to un-learn fear? A prospective paradigm to test whether children's fears can be reduced using positive information and modelling a non-anxious response. Behaviour Research and Therapy, 48(2), 164–170. [PubMed: 19875101]

Kerkhof I, Vansteenwegen D, Baeyens F, & Hermans D (2011). Counterconditioning: an effective technique for changing conditioned preferences. Experimental psychology, 58(1), 31–38. doi: 10.1027/1618-3169/a000063 [PubMed: 20382627]

Klein B (1969). Counterconditioning and fear reduction in the rat. Psychonomic Science, 17(3), 150–151.

Koizumi A, Amano K, Cortese A, Shibata K, Yoshida W, Seymour B, … Lau H (2016). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. Nature Human Behaviour, 1, 0006.

Konorski J (1967). Integrative activity of the brain: An interdisciplinary approach. Chicago: University of Chicago Press.

Konorski J, & Szwejkowska G (1956). Reciprocal transformations of heterogeneous conditioned reflexes. Acta Biologiae Experimentalis, 17, 141–165.

Kroes MC, Dunsmoor JE, Lin Q, Evans M, & Phelps EA (2017). A reminder before extinction strengthens episodic memory via reconsolidation but fails to disrupt generalized threat responses. Scientific Reports, 7(1), 10858. [PubMed: 28883499]

Krypotos A-M, & Engelhard IM (2018). Testing a novelty-based extinction procedure for the reduction of conditioned avoidance. Journal of behavior therapy and experimental psychiatry, 60, 22–28. [PubMed: 29486371]

Larrauri JA, & Schmajuk NA (2008). Attentional, associative, and configural mechanisms in extinction. Psychological review, 115(3), 640–676. doi:10.1037/0033-295x.115.3.640 [PubMed: 18729595]

LeDoux JE, Moscarello J, Sears R, & Campese V (2017). The birth, death and resurrection of avoidance: a reconceptualization of a troubled paradigm. Molecular psychiatry, 22(1), 24. [PubMed: 27752080]

LeDoux JE, & Pine DS (2016). Using neuroscience to help understand fear and anxiety: a two-system framework. American journal of psychiatry, 173(11), 1083–1093. [PubMed: 27609244]

Lee JL (2009). Reconsolidation: maintaining memory relevance. Trends in neurosciences, 32(8), 413–420. [PubMed: 19640595]

Lipp OV, & Purkis HM (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. Journal of Experimental Psychology: Animal Behavior Processes, 32(4), 431. [PubMed: 17044745]

Loerinc AG, Meuret AE, Twohig MP, Rosenfield D, Bluett EJ, & Craske MG (2015). Response rates for CBT for anxiety disorders: Need for standardized criteria. Clinical Psychology Review, 42, 72–82. [PubMed: 26319194]

Lomont JF (1965). Reciprocal inhibition or extinction? Behaviour Research and Therapy, 3(4), 209–219. [PubMed: 5858547]

Lubow RE (1973). Latent inhibition. Psychological bulletin, 79(6), 398–407. [PubMed: 4575029]

Lucas K, Luck CC, & Lipp OV (2018). Novelty-facilitated extinction and the reinstatement of conditional human fear. Behaviour Research and Therapy, 109, 68–74. [PubMed: 30120999]

Luck CC, & Lipp OV (2018). Verbal instructions targeting valence alter negative conditional stimulus evaluations (but do not affect reinstatement rates). Cognition and Emotion, 32(1), 61–80. [PubMed: 28141482]

Ludvik D, Boschen MJ, & Neumann DL (2015). Effective behavioural strategies for reducing disgust in contamination-related OCD: A review. Clinical Psychology Review, 42, 116–129. [PubMed: 26190372]

Luyten L, & Beckers T (2017). A preregistered, direct replication attempt of the retrieval-extinction effect in cued fear conditioning in rats. Neurobiology of learning and memory, 144, 208–215. [PubMed: 28765085]

Maag JW, & Kotlash J (1994). Review of Stress Inoculation Taining with Children and Adolescents: Issues and Recommendations. Behavior Modification, 18(4), 443–469. [PubMed: 7980373]

Matsumoto M, & Hikosaka O (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature, 459(7248), 837–841. [PubMed: 19448610]

McNally GP, Johansen JP, & Blair HT (2011). Placing prediction into the fear circuit. Trends Neurosci. doi:10.1016/j.tins.2011.03.005

Meichenbaum D (1985). Stress inoculation training: Pergamon Press.

Meulders A, Karsdorp PA, Claes N, & Vlaeyen JW (2015). Comparing counterconditioning and extinction as methods to reduce fear of movement-related pain. The Journal of Pain, 16(12), 1353–1365. [PubMed: 26434783]

Monfils MH, Cowansage KK, Klann E, & LeDoux JE (2009). Extinction-Reconsolidation Boundaries: Key to Persistent Attenuation of Fear Memories. Science, 324(5929), 951–955. doi:10.1126/science.1167975 [PubMed: 19342552]

Muris P, Huijding J, Mayer B, van As W, & van Alem S (2011). Reduction of verbally learned fear in children: a comparison between positive information, imagery, and a control condition. Journal of behavior therapy and experimental psychiatry, 42(2), 139–144. [PubMed: 21315873]

Nasser HM, & McNally GP (2012). Appetitive–aversive interactions in Pavlovian fear conditioning. Behavioral neuroscience, 126(3), 404. [PubMed: 22642885]

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. [PubMed: 28531872]

Norman KA, Polyn SM, Detre GJ, & Haxby JV (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in cognitive sciences, 10(9), 424–430. doi:10.1016/j.tics.2006.07.005 [PubMed: 16899397]

Paunović N (2011). Exposure inhibition therapy as a treatment for chronic posttraumatic stress disorder: A controlled pilot study. Psychology, 2(06), 605.

Pavlov IP (1927). Conditioned Reflexes. London: Oxford University Press.

Pearce JM, & Dickinson A (1975). Pavlovian countercondition: Changing the suppressive properties of shock by association with food. Journal of Experimental Psychology: Animal Behavior Processes, 1(2), 170. [PubMed: 1141821]

Pearce JM, & Hall G (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological review, 87(6), 532–552. [PubMed: 7443916]

Peck CA, & Bouton ME (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. Learning and Motivation, 21(1), 1–31. doi:10.1016/0023-9690(90)90002-6

Pedraza LK, Sierra RO, Lotz FN, & de Oliveira Alvares L (2018). Periodical reactivation under the effect of caffeine attenuates fear memory expression in rats. Scientific Reports, 8(1), 7260. [PubMed: 29740084]

Powers MB, & Emmelkamp PM (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. Journal of anxiety disorders, 22(3), 561–569. [PubMed: 17544252]

Raes AK, & De Raedt R (2012). The Effect of Counterconditioning on Evaluative Responses and Harm Expectancy in a Fear Conditioning Paradigm. Behavior Therapy, 43(4), 757–767. [PubMed: 23046778]

Ramirez S, Liu X, MacDonald CJ, Moffa A, Zhou J, Redondo RL, & Tonegawa S (2015). Activating positive memory engrams suppresses depression-like behaviour. Nature, 522(7556), 335. [PubMed: 26085274]

Reid LD (1973). Processes of fear reduction in systematic desensitization: An addendum to Wilson and Davison (1971). Psychological bulletin, 79, 107–109. [PubMed: 4684570]

Rescorla RA, & Solomon RL (1967). Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychological review, 74(3), 151. [PubMed: 5342881]

Rescorla RA, & Wagner AR (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement: Appleton-Century-Crofts.

Ressler KJ, & Mayberg HS (2007). Targeting abnormal neural circuits in mood and anxiety disorders: from the laboratory to the clinic. Nature neuroscience, 10(9), 1116. [PubMed: 17726478]

Reynolds G, Field AP, & Askew C (2018). Reductions in children's vicariously learnt avoidance and heart rate responses using positive modeling. Journal of Clinical Child & Adolescent Psychology, 47(4), 555–568. [PubMed: 27008619]

Riccio DC, Richardson R, & Ebner DL (1984). Memory retrieval deficits based upon altered contextual cues - a paradox. Psychological bulletin, 96(1), 152–165. [PubMed: 6463163]

Richardson R, Riccio DC, Jamis M, Cabosky J, & Skoczen T (1982). Modification of reactivated memory through" counterconditioning". The American journal of psychology, 67–84. [PubMed: 7125017]

Richardson R, Riccio DC, & Smoller D (1987). Counterconditioning of memory in rats. Animal Learning & Behavior, 15(3), 321–326.

Roesch MR, Esber GR, Li J, Daw ND, & Schoenbaum G (2012). Surprise! Neural correlates of Pearce–Hall and Rescorla–Wagner coexist within the brain. European Journal of Neuroscience, 35(7), 1190–1200. [PubMed: 22487047]

Salinas-Hernández XI, Vogel P, Betz S, Kalisch R, Sigurdsson T, & Duvarci S (2018). Dopamine neurons drive fear extinction learning by signaling the omission of expected aversive outcomes. eLife, 7, e38818. [PubMed: 30421719]

Scavio MJ, & Gormezano I (1980). Classical-classical transfer: Effects of prior appetitive conditioning upon aversive conditioning in rabbits. Animal Learning & Behavior, 8(2), 218–224.

Schweckendiek J, Klucken T, Merz CJ, Kagerer S, Walter B, Vaitl D, & Stark R (2013). Learning to like disgust: neuronal correlates of counterconditioning. Frontiers in human neuroscience, 7, 346. [PubMed: 23847514]

Serino S, Triberti S, Villani D, Cipresso P, Gaggioli A, & Riva G (2014). Toward a validation of cyber-interventions for stress disorders based on stress inoculation training: a systematic review. Virtual Reality, 18(1), 73–87.

Sharp S, Thomas C, Rosenberg L, Rosenberg M, & Meyer III W (2010). Propranolol does not reduce risk for acute stress disorder in pediatric burn trauma. Journal of Trauma and Acute Care Surgery, 68(1), 193–197.

Shiban Y, Pauli P, & Muhlberger A (2013). Effect of multiple context exposure on renewal in spider phobia. Behaviour Research and Therapy, 51(2), 68–74. doi:10.1016/j.brat.2012.10.007 [PubMed: 23261707]

Sitaram R, Ros T, Stoeckel L, Haller S, Scharnowski F, Lewis-Peacock J, … Sulzer J (2017). Closed-loop brain training: the science of neurofeedback. Nature Reviews Neuroscience, 18(2), 86. [PubMed: 28003656]

Skinner BF (1938). The Behavior of Organisms: An Experimental Analysis. New York and London: Appleton-Century Company.

Skinner MD, Lahmek P, Pham H, & Aubin H-J (2014). Disulfiram efficacy in the treatment of alcohol dependence: a meta-analysis. PloS one, 9(2), e87366. [PubMed: 24520330]

Solomon RL, & Corbit JD (1978). An opponent-process theory of motivation. The American Economic Review, 12–24.

Taschereau-Dumouchel V, Cortese A, Chiba T, Knotts J, Kawato M, & Lau H (2018). Towards an unconscious neural reinforcement intervention for common fears. Proceedings of the National Academy of Sciences, 115(13), 3470–3475.

Thomas BL, Cutler M, & Novak C (2012). A modified counterconditioning procedure prevents the renewal of conditioned fear in rats. Learning and Motivation, 43(1–2), 24–34. doi:10.1016/j.lmot.2012.01.001

Treanor M, Brown LA, Rissman J, & Craske MG (2017). Can memories of traumatic experiences or addiction be erased or modified? A critical review of research on the disruption of memory reconsolidation and its applications. Perspectives on Psychological Science, 12(2), 290–305. [PubMed: 28346121]

Tryon WW (2005). Possible mechanisms for why desensitization and exposure therapy work. Clinical Psychology Review, 25(1), 67–95. [PubMed: 15596081]

Tye KM, Stuber GD, de Ridder B, Bonci A, & Janak PH (2008). Rapid strengthening of thalamoamygdala synapses mediates cue–reward learning. Nature, 453(7199), 1253. [PubMed: 18469802]

van Dis EA, Hagenaars MA, Bockting CL, & Engelhard IM (2019). Reducing negative stimulus valence does not attenuate the return of fear: Two counterconditioning experiments. Behaviour Research and Therapy, 103416. [PubMed: 31254717]

Van Gucht D, Baeyens F, Vansteenwegen D, Hermans D, & Beckers T (2010). Counterconditioning reduces cue-induced craving and actual cue-elicited consumption. Emotion, 10(5), 688. [PubMed: 21038951]

Watson JB, & Rayner R (1920). Conditioned emotional reactions. Journal of Experimental Psychology, 3, 1–14.

Wiederhold BK, & Wiederhold MD (2008). Virtual reality for posttraumatic stress disorder and stress inoculation training. Journal of CyberTherapy & Rehabilitation, 1(1), 23–35.

Wilson EH, & Dinsmoor JA (1970). Effect of feeding on" fear" as measured by passive avoidance in rats. Journal of comparative and physiological psychology, 70(3p1), 431. [PubMed: 5418469]

Wilson GT (1973). Counterconditioning versus forced exposure in extinction of avoidance responding and conditioned fear in rats. Journal of comparative and physiological psychology, 82(1), 105. [PubMed: 4684966]

Wilson GT, & Davison GC (1971). Processes of fear reduction in systematic desensitization: Animal studies. Psychological bulletin, 76(1), 1. [PubMed: 5105019]

Wolpe J (1954). Reciprocal inhibition as the main basis of psychotherapeutic effects. AMA Archives of Neurology & Psychiatry, 72(2), 205–226. [PubMed: 13180056]

Wolpe J (1961). The systematic desensitization treatment of neuroses. The Journal of nervous and mental disease, 132, 189–203. doi:10.1097/00005053-196103000-00001 [PubMed: 13786444]

Wolpe J, & Plaud JJ (1997). Pavlov's contributions to behavior therapy: The obvious and the not so obvious. American Psychologist, 52(9), 966. [PubMed: 9382243]

Wood NE, Rosasco ML, Suris AM, Spring JD, Marin MF, Lasko NB, … Pitman RK (2015). Pharmacological blockade of memory reconsolidation in posttraumatic stress disorder: Three negative psychophysiological studies. Psychiatry Research, 225(1–2), 31–39. doi:10.1016/j.psychres.2014.09.005 [PubMed: 25441015]

Woods AM, & Bouton ME (2007). Occasional reinforced responses during extinction can slow the rate of reacquisition of an operant response. Learning and Motivation, 38(1), 56–74. [PubMed: 19132143]

Zbozinek TD, Hermans D, Prenoveau JM, Liao B, & Craske MG (2015). Post-extinction conditional stimulus valence predicts reinstatement fear: Relevance for long-term outcomes of exposure therapy. Cognition and Emotion, 29(4), 654–667. [PubMed: 24957680]

Highlights

- Counterconditioning is a technique and process of behavior modification
- Whether counterconditioning is more or less effective than extinction is unclear
- We review behavioral and neural findings of counterconditioning across species
- We hope this review motivates further research on this underexplored topic

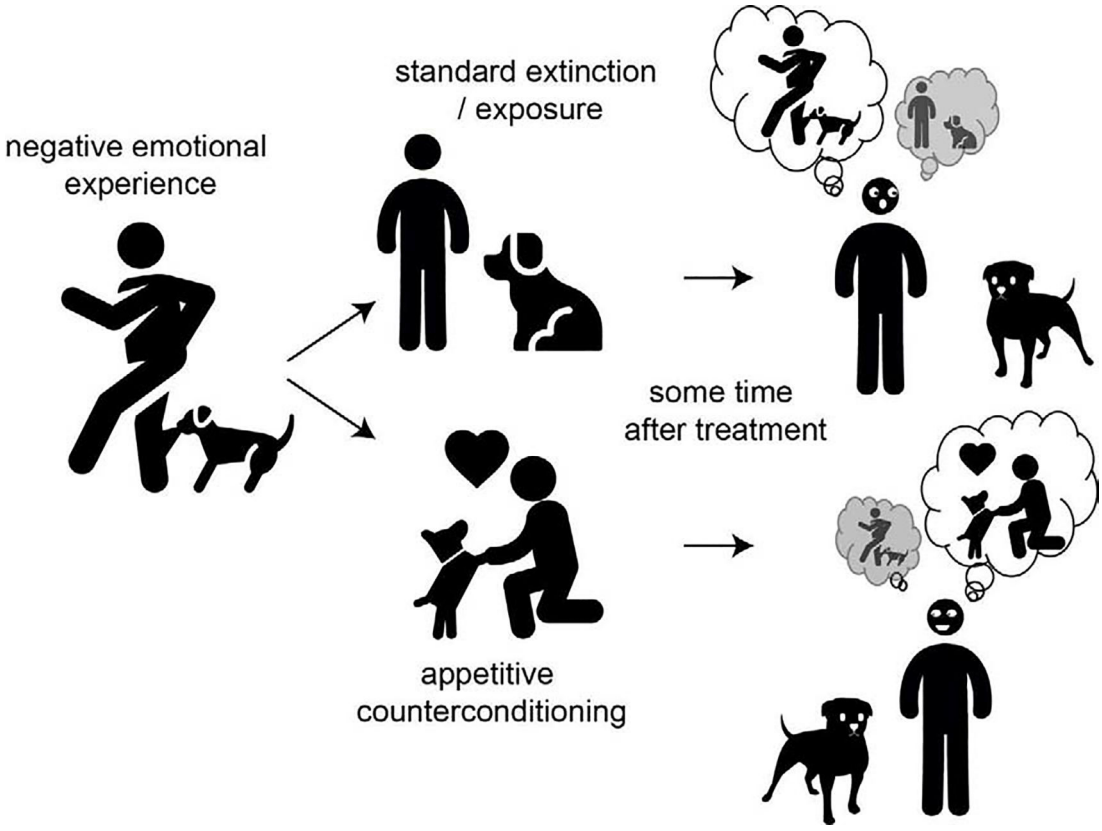

Figure 1.

Counterconditioning, as distinct from standard extinction, involves replacing an expected salient outcome with a new outcome of the opposite valence. Counterconditioning as a technique formed the basis for therapies such as systematic desensitization or aversion therapy, where unwanted responses are inhibited by activating an antagonistic response system. In this simplified illustration, a person afraid of dogs after an encounter with a vicious dog, might undergo extinction-based treatment via exposure to the feared stimulus in a safe environment. Alternatively, a counterconditioning approach could involve gradually encountering dogs while in a relaxed or pleasant state, or while performing an appetitive action that inhibits expression of fear. Both approaches are considered forms of new learning that generate a second association (e.g., dogs are safe), which later competes for expression with the original association (e.g., dogs are dangerous). In this illustration, a patient shows relapse to the original association following extinction but not counterconditioning. However, laboratory research on the long term effects of counterconditioning versus extinction shows varied results.

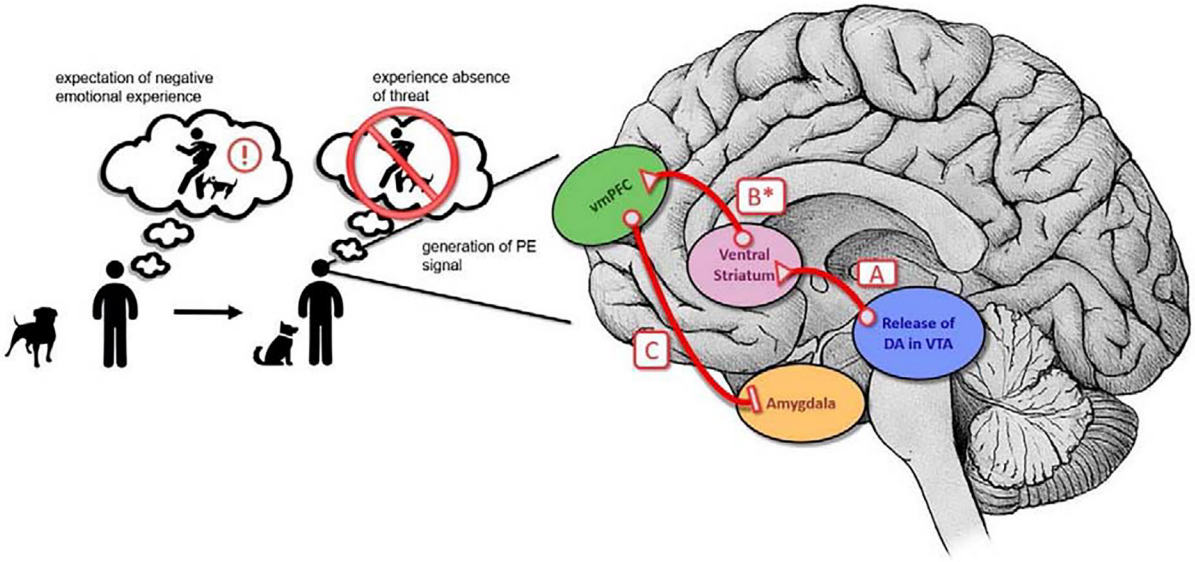

Figure 2. A simplified dopaminergic mediated fear extinction circuit.

When an aversive outcome is omitted during extinction (or exposure), a prediction error (PE) is generated by the dopamine system. (A) The PE is encoded by a subpopulation of dopamine neurons in the ventral tegmental area (VTA) (Salinas-Hernandez et al., 2018; Luo et al., 2016). This subclass of dopamine neurons transmits the PE from the VTA to the ventral striatum. (B) VTA dopaminergic neurons projecting to the ventral striatum promote the activity of extinction related circuitry, particularly the ventromedial prefrontal cortex (vmPFC) and amygdala (Luo et al., 2018; Correia et al., 2016; Rodriguez-Romageura et al., 2012). (*) Even though these studies support the role of dopamine in extinction, it is still not entirely clear how dopamine in the ventral striatum modulates activity in the medial prefrontal cortex and amygdala. (C) The vmPFC regulates the expression of learned fear by gating output from the central nucleus of the amygdala.

Table 1.

Literature overview of counterconditioning studies in humans.

A list of contemporary studies in humans comparing the effectiveness of counterconditioning vs standard extinction. This list is meant to provide an overview of the field, but is not an exhaustive list. Behaviors listed under "CC > EXT" are those specifically affected by counterconditioning, whereas behaviors listed under "CC = EXT" were equivalent between conditions. Comparison between counterconditioning and extinction condition does not verify strength of the statistical test used. CC = counterconditioning; EXT = standard extinction; CS= conditioned stimulus; US = unconditioned stimulus; SCR = skin conductance responses; APT = affective priming test.

						Behavior affected by counterconditioning versus extinction	**Behavior unaffected by counterconditioned relative to extinction**
Publication	**CC Paradigm**	**CS**	**US1**	**US2**	**Betweensession followup?**		
Dunsmoor et al. 2015	aversive-toneutral	angry face	shock	tone	✓	SCR	
Dunsmoor et al. 2019	aversive-toneutral	angry face	shock	tone	✓	SCR	
Lucas et al. 2018	aversive-toneutral	angry face	shock	tone	✓	SCR	
Krypotos & Englehard	aversive-toneutral	angry face	shock	tone	✓		avoidance
Raes and Raedt, 2012	aversive-toneutral	neutral face	100 decibel white noise	neutral tone		CS APT scores	US1 expectancy, CS valence, CS fear ratings
Raes and Raedt, 2012	aversive-toappetitive	neutral face	100 decibel white noise	baby laugh		CS APT scores	US1 expectancy, CS valence, CS fear ratings
Kang et al. 2018	aversive-toappetitive	spider image	shock	cartoon images	✓	US1 expectancy ratings	CS valence ratings
van Dis et al. 2019	aversive-toappetitive	neutral face	shock	6s funny film clips	✓	CS valence ratings	US1 expectancy, SCR & FPS
Meulders et al. 2015	aversive-toappetitive	joystick movement	painful shock	monetary reward		CS valence (borderline significance)	US1 fear, US1 expectancy, US1 pain ratings
de Jong et al. 2000	aversive-toappetitive	spider	participants were women with spider phobia	tasty food items & favorite music	✓		affective valence of spiders
Eifert et al. 1988	aversive-toappetitive	feared animals	participants were patients with animal phobia	positively rated music	✓	affective valence of feared animals	
Dour et al. 2016	aversive-toappetitive (valence training after exposure)	spiders	participants had elevated fear of spiders	7 minute positive film clip about spiders	✓	CS fear ratings, CS avoidance, CS valence	

						Behavior affected by counterconditioning versus extinction	Behavior unaffected by counterconditioned relative to extinction
Publication	**CC Paradigm**	**CS**	**US1**	**US2**	**Betweensession followup?**		
Luck & Lipp 2017	aversive-toappetitive	image of a person	shock	positive information about the CS		CS valence	SCR
Newall et al., 2017	aversive-toappetitive	images of novel animals	fearful adult face	happy adult face		CS avoidance and fear ratings	
Reynolds et al., 2017	aversive-toappetitive	images of animals	fearful adult face	happy adult face		CS avoidance fear ratings & heart rate	
Kelly et al., 2010	aversive-toappetitive	two unfamiliar Australian marsupials	threat information about the animal	positive information about the animal		CS fear beliefs	avoidance
Muris et al., 2011	aversive-toappetitive	novel animal	threat information about the animal	positive information about the animal		CS fear beliefs	
Van Gucht et al., 2010	appetitiveto-aversive	serving tray	chocolate	polysorbate 20	✓	US1 expectancy and liking	
Kaag et al., 2016	appetitiveto-aversive	colored squares	monetary reward	shock		reduced activity in reward network	