Neuropsychopharmacology (2014) 39, 515–527

www.neuropsychopharmacology.org

Review

Learning not to Fear: Neural Correlates of Learned Safety

Eryan Kong[1], Francisco J Monje[1], Joy Hirsch[2,3,4,5] and Daniela D Pollak*[,1]

[1]Department of Neurophysiology and Neuropharmacology, Center for Physiology and Pharmacology, Medical University of Vienna, Vienna, Austria; [2]Department of Neuroscience, Columbia University, New York, NY, USA; [3]fMRI Research Center, Columbia University, New York, NY, USA; [4]Department of Radiology, Columbia University, New York, NY, USA; [5]Department of Psychology, Columbia University, New York, NY, USA

The ability to recognize and properly respond to instances of protection from impending danger is critical for preventing chronic stress and anxiety—central symptoms of anxiety and affective disorders afflicting large populations of people. Learned safety encompasses learning processes, which lead to the identification of episodes of security and regulation of fear responses. On the basis of insights into the neural circuitry and molecular mechanisms involved in learned safety in mice and humans, we describe learned safety as a tool for understanding neural mechanisms involved in the pathomechanisms of specific affective disorders. This review summarizes our current knowledge on the neurobiological underpinnings of learned safety and discusses potential applications in basic and translational neurosciences.
Neuropsychopharmacology (2014) **39,** 515–527; doi:10.1038/npp.2013.191; published online 18 September 2013

Keywords: learned safety; learned fear; conditioned inhibition; amygdale; anxiety; depression

INTRODUCTION

Awareness and appropriate response to imminent threats are essential for one's well-being and self-preservation. The physical and emotional reactions initiated in response to such menaces are commonly termed as fear. Complementing instinctive (innate) and acquired (learned) fear, an alternative set of physiological responses—innate and learned safety, respectively—is triggered by the search for and identification of circumstances that provide protection from impending danger. The concept of learned behavioral responses has evolved from the seminal studies of Ivan P Pavlov in the 1920s, in which he discovered associative learning, whereby an *a priori* neutral signal, the conditioned stimulus (CS) such as a bell, becomes a predictor of an inherently relevant stimulus, the unconditioned excitatory stimulus (US) such as food. The associative learning process is based on repetitions of combined presentations of the CS and the US through which the CS develops the ability to elicit the behavioral response that was originally induced by the US, such as saliva secretion (excitation). Pavlov described this association as 'conditioned reflexes' (Pavlov, 1927). Further, he described 'conditioned inhibition', whereby a different stimulus (the CS−), which is never accompanied by the US (food serving) in a series of training trials, becomes an inhibitor of the excitatory behavioral response (saliva secretion), previously elicited by the excitatory CS (CS+).

*Correspondence: Professor DD Pollak, Department of Neurophysiology and Neuropharmacology, Center for Physiology and Pharmacology, Medical University of Vienna, Schwarzspanierstrasse 17, Vienna 1090, Austria, Tel: +43 1 40160 32170, Fax: +43 1 40160931201, E-mail: daniela.pollak@meduniwien.ac.at
Received 14 March 2013; revised 5 July 2013; accepted 13 July 2013; accepted article preview online 21 August 2013

Robert Rescorla (1969, 1971) later defined Pavlovian's excitation and inhibition as opposite associative processes, a view that quickly found wide acceptance (Williams *et al*, 1992). Most importantly, Rescorla set up two specific behavioral tests, called summation and retardation tests, to evaluate the status of a CS as inhibitory. In the summation test, a CS− weakens the behavior evoked by a CS+ when the two stimuli are jointly introduced to an animal, whereas in the retardation test, a CS− acquires the properties of an excitatory stimulus slower when coupled with US rather than a neutral stimulus (Rescorla, 1971).

A special case of conditioned inhibition, which is capable of reducing the behavioral responses evoked by conditioned fear (also called learned fear), is conditioned safety (also called learned safety) (Rogan *et al*, 2005). Learned safety and learned fear are opposite associative processes that are important for survival and well-being (Pollak *et al*, 2008, 2010b). In humans, pathological forms of learned fear are hallmarks of severe psychopathologies, such as anxiety disorders, post-traumatic stress disorders (PTSD), and depression (Pollak *et al*, 2008). The potential therapeutic application of learned safety in patients suffering from fear-related emotional disturbances is notable, as it modulates behavioral responses and neural circuitries induced by learned fear in both mice and humans (Pollak *et al*, 2008).

Acquired fear responses, which can be studied in laboratory animals using the paradigm of fear conditioning, have a prominent role in various psychiatric conditions. Insights into these underlying neurobiological mechanisms have been previously reviewed, and thus, will not be the focus of this review. Advances in the development and in-depth characterization of an animal model for inhibition of fear responses may enhance our understanding of the systemic, cellular, and molecular processes engaged during

learned safety and may increase the translational potential for the investigation of learned safety.

This review summarizes the latest developments in the field of learned safety research by focusing on addressing the following major points: How does learned safety compare with learned fear and other related behavioral paradigms? Which brain regions and neural circuits are involved in this learning process? The cellular and molecular mechanisms of safety learning, as well as its translational potential, are also discussed.

HOW DOES LEARNED SAFETY RELATE TO LEARNED FEAR AND OTHER BEHAVIORAL PARADIGMS OF EMOTIONAL LEARNING?

Learned Safety and Learned Fear

Learned fear and learned safety are both associative learning paradigms. Learned fear results from a positive correlation of an intrinsically neutral CS and an intrinsically aversive US, resulting in the ability of the neutral CS to predict an aversive event. By contrast, learned safety results from a negative correlation of a neutral CS and an aversive US. In such circumstances, the neutral CS develops the ability to predict protection from the aversive US, in the event of a safe situation. Thus, the CS acts to inhibit learned fear responses, a behavioral phenomenon called conditioned inhibition (of learned fear) (Pollak *et al*, 2008) (Figure 1). Different experimental approaches have been used to induce learned safety in rodents and humans, with successful safety learning being evaluated both at the behavioral and neural activity level (see Table 1 for some selected paradigms). In mice and rats, explicit unpaired procedures, together with shock off-set pairings, conditional discrimination paradigms, and active avoidance procedures are mainly being used to study the behavioral, cellular, and molecular correlates of learned safety (Figure 2 and Table 1). In humans, conditional discrimination paradigms as well as protocols of explicit unpaired presentations of US and CS are being used for studying the behavioral and neural effects of learned safety as well as to investigate differential responses in patients with psychiatric disorders (Figure 3 and Table 1).

The effect of learned safety is not limited to its effect on conditioned fear, rather it has also been shown to independently have rewarding properties and facilitating effects on behavior (Ganguly and Kleinfeld, 2004; Masuda *et al*, 1994; Rogan *et al*, 2005), indicating that learned safety may be associated with positive affective brain states. Indeed, learned safety can be thought of as a learning process by which animals acquire the ability to take advantage of sources of safety and security in the environment (Pollak *et al*, 2008), escape from an aversive or dangerous situation, and, therefore, to experience relief from the ongoing stress of imminently threatening conditions (Pollak *et al*, 2008). Hence, learned fear predicts an upcoming dangerous situation and learned safety inhibits the physiological responses evoked by learned fear and presents an active mechanism for the identification of protection, like the shelter of a nest, in the surrounding environment (Rogan *et al*, 2005). Therefore, as numerous different paradigms for the induction of learned fear and consequently several behavioral readouts exist, also learned safety can be studied in experimental animal using independent procedures and varying behavioral displays for the valuation of its effects (see two exemplary protocols depicted in Figure 2). Therefore, learned safety and learned fear can be expected to be served by related, but independent neural circuitries. This notion is also supported by the observation that although learned safety acts as a behavioral antidepressant reducing immobility, no effect of the learned fear signal on depression-like behavior in the forced-swim test has been described (Pollak *et al*, 2008). This finding presumably reflects, in an applied sense, the early dictum that 'Conditioned inhibition is not the symmetrical opposite of conditioned excitation' (Baker, 1974). As such, while the effects of a fear CS may,

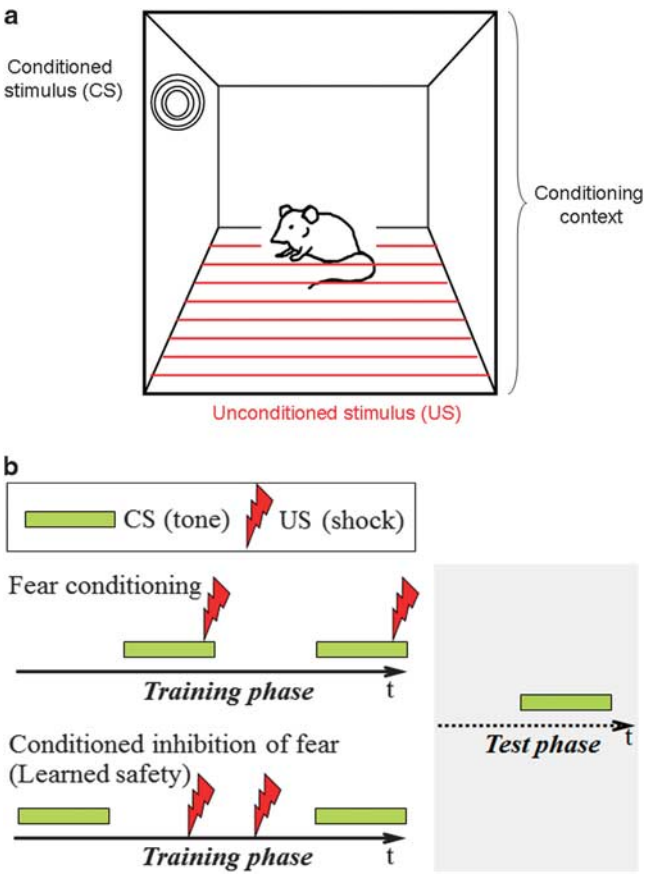

Figure 1 Fear conditioning and conditioned inhibition of fear are based upon associative learning process involving the conditioned stimulus (CS), the unconditioned stimulus, and conditioning context. (a) Classical conditioning paradigms based on associative learning involve the conditioning chamber where the training procedures are carried out which constitutes the conditioning context. During training, the unconditioned stimulus (US), which has an intrinsic valence (here, aversive such as a mild electric footshock) becomes associated with the CS (such as an auditory signal) that is *a priori* neutral. (b) During fear conditioning, the temporal pairing of the US and the CS induces a transfer of the fear-inducing properties from the US to the CS. Consecutively, the previously neutral stimulus CS and the conditioning context elicit the physiological and behavioral responses (such as freezing) inherent to the US. Conditioned inhibition of fear (or learned safety) is mediated by the temporal dissociations of the US and the CS, in a way that the two stimuli never coincide. Consequently, the presence of the conditioned inhibitor leads to a reduction of the fear response induced by the conditioning context.

Table 1 Procedural Characteristics and Key Findings of Learned Safety Studies in Humans and Rodents

Paradigm	Behavioral/ physiological tests	Subject/strain	Key finding	Refs
Human				
Conditional discrimination	Fear-potentiated startle	Healthy volunteers	Safety signal reduces anticipatory anxiety to threat stimulus	Grillon *et al* (1994b)
		PD patients	Deficient discriminative learning to learned safety and danger cues driven by enhanced startle potentiation to the learned safety cue	Lissek *et al* (2009)
		PTSD and MDD patients	Absence of fear inhibition to safety cues	Jovanovic *et al* (2010)
	Skin conductance	Healthy volunteers	Dissociation within the ventromedial PFC between a safe stimulus previously predicting danger and a 'naive' safe stimulus	Schiller *et al* (2008)
Explicit unpairing CS–US	Pupillary diameter	Healthy volunteers	Learned safety involves reduced amygdalar and heightened dorsolateral PFC neural activity	Pollak *et al* (2010b)
Rodent				
Explicit unpairing CS–US	Fear-potentiated startle	Sprague–Dawley rats	Neural dissociations between the processing of appetitive and safety signals exist	Josselyn *et al* (2005)
	Summation and retardation (freezing)		Learned safety leads to a reduction in spine size on synapses of the LA	Ostroff *et al* (2010)
	Summation and retardation test (freezing), Open field, place preference	C57BL/6J mice	Learned safety reduces learned and instinctive fear, as well as positive affective responses	Rogan and LeDoux (1995)
	Summation and retardation test (freezing), FST, SPT	C57BL/6N mice	Learned safety acts as a behavioral antidepressant	Pollak *et al*(2008)
	Summation and retardation (freezing)	C57BL/6J, 129S1/SvImJ (S1) mice	S1 mice exhibit deficiencies in safety learning	Ostroff *et al* (2010)
Inescapable shock (off-set pairing)	Social exploration	Sprague–Dawley rats	The sensory insula has a critical role in learned safety	Christianson *et al* (2008)
			Safety signal-mediated reduction in neural fear responses during uncontrollable stressors involves the sensory insular cortex and BNST	Christianson *et al* (2011)
Active avoidance	Summation and retardation test (suppression of licking)	Wistar rats	Safety signal behaves as a conditioned inhibitor after long avoidance procedure	Cándido *et al* (2004)
Infant odor–shock pairing	FST, SPT	Long–Evans rats	An odor, which acquired characteristics of the maternal odor, serves as safety signal to revert depressive-like behavior and amygdala activity in adulthood, even when paired with shock infancy	Sevelinges *et al* (2011)

Abbreviations: BNST, bed nucleus stria terminalis; FST, forced-swim test; LA, lateral amygdala; MDD, major depressive disorder; PFC, prefrontal cortex; PD, panic disorder; PTSD, post-traumatic stress disorder; SPT, sucrose preference test.

determined by its neurobiological underpinnings of how and where negative emotional memories are formed and stored in the brain, specifically relate to the induction of a conditioned fear response, the neural circuitry and molecular mechanisms subserving learned safety may allow for more behavioral flexibility, leading to the applicability of the safety signal independent of the specific conditions under which the safety response had originally been acquired. However, it is known that overtraining protocols of fear learning also lead to a generalization of the fear response, where a defensive response is elicited even in environmental context or during exposure to cues distinct to the conditioning setting (Laxmi *et al*, 2003). Interestingly, enhanced generalization of fear conditioning has been observed in juvenile mice as compared with adult mice (Ito *et al*, 2009). This generalization, however, was significantly reduced in the presence of the explicitly unpaired cue (ie, the safety signal) in both juvenile and adult animals, suggesting that the neurobiological mechanisms required for safety learning are already functional during the adolescent period and have adaptive relevance for overcoming augmented juvenile fear generalization (Ito *et al*, 2009). In light of these findings, it would be tempting to explore whether, because of this heightened emotionality in adolescent animals, an

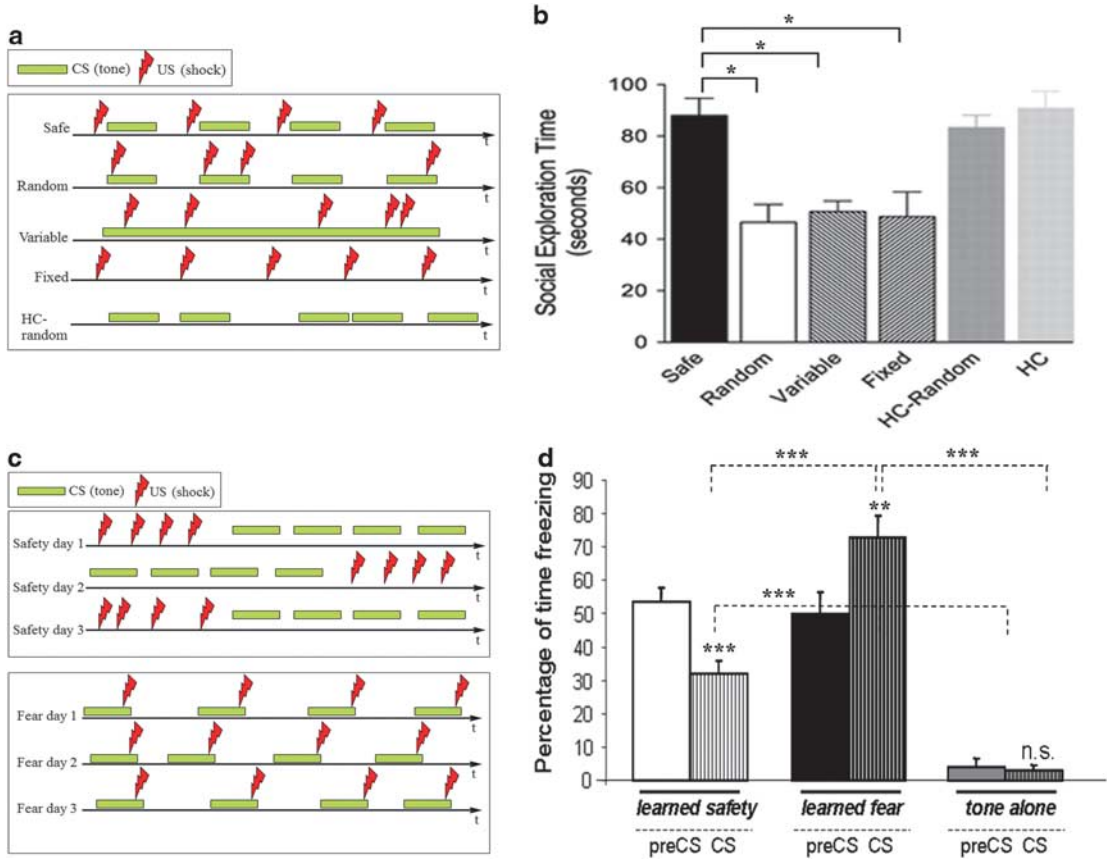

Figure 2 Behavioral paradigms and readouts of learned safety in rodents. (a and b) Inescapable shock (off-set pairing) procedure to induce learned safety in rats. (a) Schematic illustration of the different shock and safety signal conditions used. Red filled bars represent the occurrence of tail shock, and green filled boxes indicate the occurrence of a safety CS (5 s chamber blackout) over time. In the Safe group, the CS was presented with the termination of each shock US. In the Random group, the CS was delivered independent of the shock schedule. In the variable group, shocks were delivered at the same schedule as in the Random group, but lights remained on throughout the session. Animals in the Fixed group received 100, 5 s shocks with the house light on throughout the session. Rats in the home cage control group (HC) were left undisturbed in their home cages and animals in HC-Random group remained in the home cage but were exposed to 100, 5 s blackouts in a room adjacent to the stress room. (b). Mean (±SEM) time spent exploring the juvenile conspecific in a 3 min test given 24 h after 100 tail shocks (Christianson *et al*, 2008). Group designations indicate the conditions of previous tail shocks. Pairwise comparisons identified significant differences between Safe and all other groups receiving shock. (c and d) Explicit unpairing procedure to induce learned safety in mice. (c) Schematic illustration of safety and fear conditioning used. Safety conditioning (upper panel) consisted of a simple conditioned inhibition of learned fear paradigm in which the delivery of four shock US is followed by the presentation of four tone CS. In the fear conditioning protocol (bottom panel), the number of CS and US presentations was matched to the safety conditioning paradigm (ie four paired CS–US). Training was conducted over a period of 3 days, one session per day. A memory recall test, consisting of a single CS presentation, was carried out 24 h after the last training day. (d) Contextual freezing in the presence of the CS in safety conditioned, fear conditioned, and tone alone control mice. $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$. Reproduced, with permission, from Christianson *et al* (2008) and Pollak *et al* (2008, 2010a).

intense fear conditioning procedure could also lead to an induction of immobility in the forced-swim test.

When contrasting learned safety and learned fear, several important insights about the neurobiology of learned safety can also be drawn from a host of studies in which explicitly unpaired 'control' groups had been studied. Important examples include the assessment of generalization described above (Bang *et al*, 2008; Ito *et al*, 2009; Laxmi *et al*, 2003), the effect on sleep (Jha *et al*, 2005; Madan *et al*, 2008), and investigations on the role of the amygdala (Choi *et al*, 2001; Maren, 2000; Maren *et al*, 2001).

Learned Safety and Fear Extinction

Related to but distinct from learned safety is another behavioral paradigm: fear extinction (ie, the extinction of learned fear). Fear extinction is an active relearning process in which animals are first trained in a fear conditioning protocol, and then with a series of unreinforced presentations of the CS alone (no longer paired with the aversive event) that leads to a reduction of conditioned fear response (Kaplan and Moore, 2011; Lattal and Maughan, 2012). New learning involved in fear extinction is in fact proposed to be a suppression of the expression of conditioned fear rather than the removal of the original fear memory (Bouton *et al*, 2006; Jungling *et al*, 2008; Maren and Quirk, 2004; Myers and Davis, 2007). Thus, both fear extinction and learned safety lead to the inhibition of responses evoked by fear, yet the approach and the involved behavioral repertoire may be distinct. Fear extinction requires pretraining of fear conditioning, whereas learned safety involves prior learning that the CS never coincides with an aversive event by a series of unpaired presentation of the two stimuli. Moreover, while the behavioral effects of fear extinction are

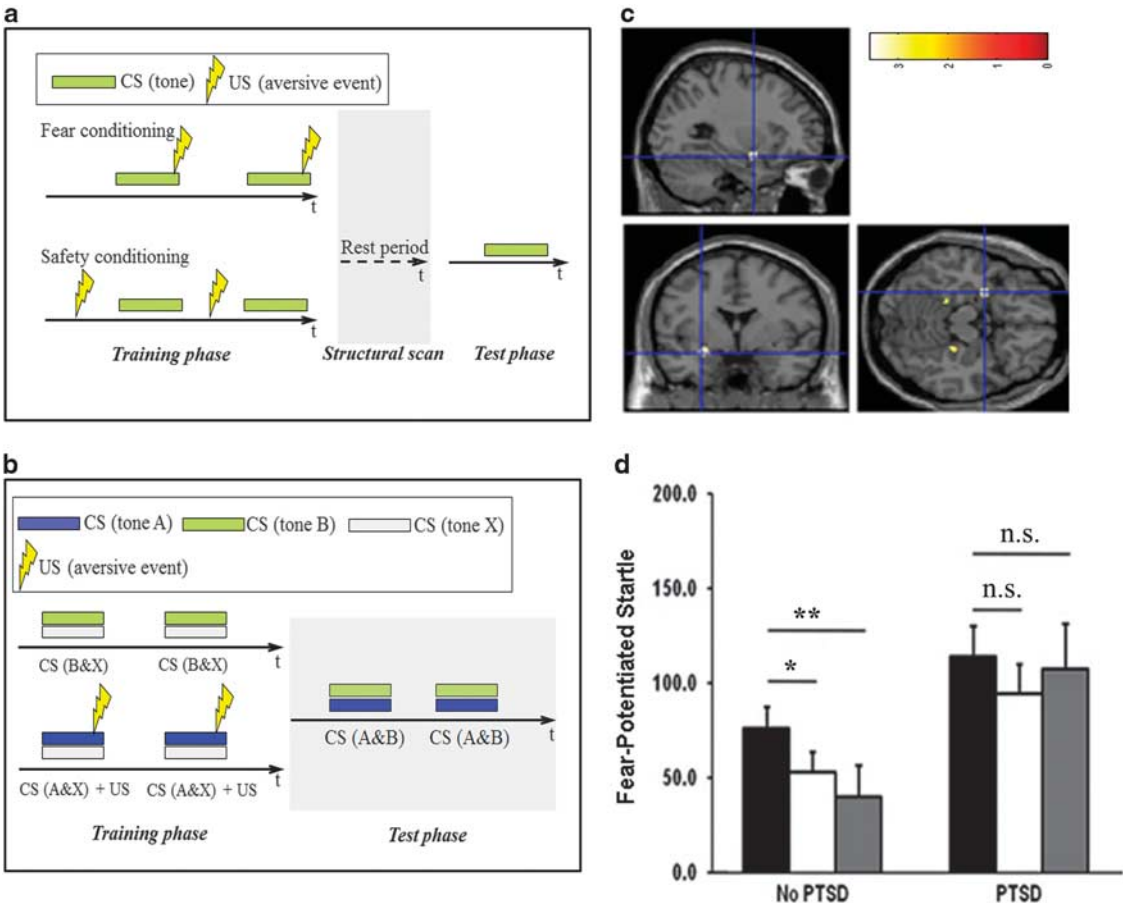

Figure 3 Human paradigms and functional consequences of learned safety. (a) An explicit unpairing protocol: experimental paradigm and amygdala responses to the aversive stimulus. training and test phase: The training phase (left) consisted of several explicitly unpaired (bottom row) or paired (top row) presentations of the conditioned stimulus (CS) and the unconditioned stimulus (US) and was followed by a period of rest (middle) during which the structural MRI (magnetic resonance imaging) images were acquired. The test phase (right) consisted of five presentations of the CS alone (Pollak *et al*, 2010b). (b) A conditional discrimination procedure: diagram of the trial design in the AX + /BX − human paradigm. The training phase (left) consisted of four unpaired CS (B and X) alone (bottom row) and paired CS (A and X) presentations and the US (top row). The test phase (right) consisted of three presentations of the CS (A and B) (Jovanovic *et al*, 2010). (c) A cluster of differential activation in the left amygdala between safety and fear trained subjects in response to the CS is shown on a standard brain. Color codes indicate the *t* score = 3.56 days. Mean fear-potentiated startle on AX +, BX −, and AB trials across diagnostic groups from three studies. Fear-potentiated startle in a traumatized civilian sample with post-traumatic stress disorders (PTSD) ($n = 29$) and without PTSD ($n = 61$) (Jovanovic *et al*, 2010). Reproduced, with permission, from Jovanovic *et al* (2010) and Pollak *et al* (2010b) (b) and (d).

limited to the inhibition of the learned fear induced by a specific CS, learned safety involves a wider spectrum of behavioral responses, including the reduction of innate fear, the potential of the stimulus to be transferred and to elicit reward- and antidepressant-like effects (Pollak *et al*, 2008; Rogan *et al*, 2005; Sevelinges *et al*, 2011). Interestingly, however, both paradigms have an important role as animal models of PTSD, by reproducing separate symptomatic features forming part of the clinical diagnostic picture of the disorder and presumably reflecting distinct endophenotypes (see below for detailed discussion).

THE NEURAL CIRCUITRY OF LEARNED SAFETY

Safety learning and memory of learned safety most likely involve the consorted action of a network of brain regions, each of which is recruited to exert a particular functional role determined by its neuronal integration and the specific stage of safety learning (ie, acquisition during conditioning, consolidation of the learned information, and stabilization of the memory and recall together with the behavioral expression of learned safety in response to the safety signal). A simplified working model of the neural circuitry proposed to be mediating learned safety is presented in a schematic illustration in Figure 4.

The Role of the Amygdala and the Striatum

Considering the role of the amygdala as a site for integration of the CS and the US during fear conditioning, it appeared as obvious target for the first attempts to delineate the neural circuitries underlying learned safety. An electrophysiological approach demonstrated related but contrasting neural signatures of learned safety in comparison to learned fear in the lateral amygdala (LA) and the caudoputamen (CP) (Rogan *et al*, 2005). Although learned safety leads to a decrease in slope and amplitude of the CS-evoked field potential in LA, learned fear induces an increase in the CS-evoked field potential (Rogan *et al*, 1997).

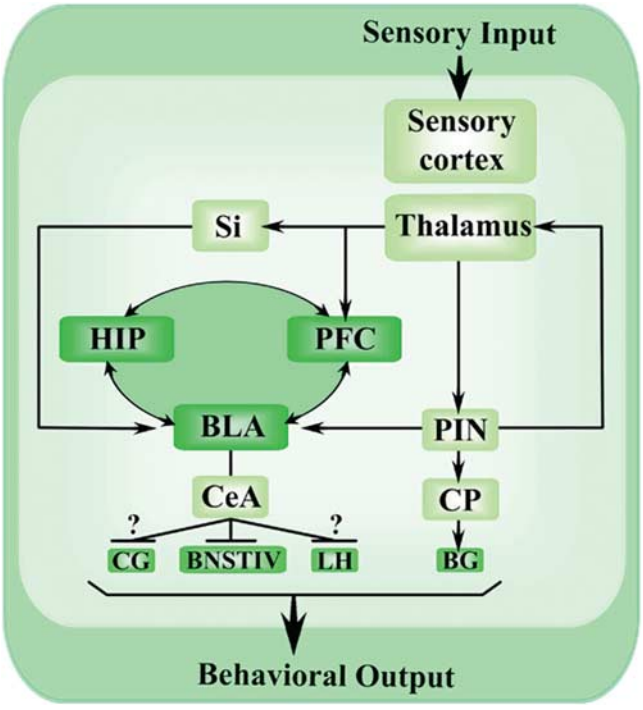

Figure 4 Model for the potential neural circuitry mediating learned safety. In such a model, which is largely based on rodent studies, the sensory insular (Si) and posterior intralaminar nucleus (PIN) work in concert with the basolateral amygdala (BLA), leading to the inhibition of bed nucleus of the striatum terminals (BNSTIv) and possibly also to other output regions of amygdala (ie, the central gray (CG) and the lateral hypothalamus (LH)) to mediate the behavioral effects of learned safety. The sensory input of the signal used to induce learned safety is received by the thalamus and the sensory cortex, presumably also receiving direct sensory inputs, which project to the PIN and the Si. The Si projects directly to the BLA, which also receives input from the PIN and orchestrates the behavioral output through communications with the central amygdala (CeA). The PIN, furthermore, projects back to the thalamus and also transmits signals to the part of the caudoputamen (CP) lying dorsal to BLA, which may contribute to the emotional regulation of the behavioral output through its connection to the basal ganglia (BG). Cortical control mechanisms are thought to be mediated by the prefrontal cortex (PFC) through direct inhibitory constraints on the BLA but also by its interaction with the hippocampus (HIPP) required for gating the modulatory influences of prefrontal regions, hereby leading to an inhibition of the emotional response orchestrated by the amygdala during learned safety.

However, in the CP, learned safety markedly increases the slope and amplitude of CS-evoked field potential, whereas no effect of learned fear was observed (Rogan *et al*, 2005). Moreover, a reduction of stressor-induced c-Fos-immunoreactive cells, indicating neural activity induced by safety signals, was also observed in another study on basolateral amygdala (BLA), which also reports a similar effect of learned safety in the ventrolateral region of the bed nucleus of stria terminalis (BNSTlv) (Christianson *et al*, 2008; Christianson *et al*, 2011). Furthermore, deficient paired-pulse inhibition in the amygdala and the piriform cortex induced by infant odor-shock pairing is restored in the presence of the infant odor functioning as a safety signal (Sevelinges *et al*, 2011). Thus, the amygdala most likely constitutes the prime site for both acquisition and consolidation of learned safety subserved by specific molecular events reflected in safety learning-induced gene expression in the BLA (discussed below).

In a translational approach, the involvement of the amygdala and the striatum in learned safety were confirmed in a functional neuroimaging study in humans, where exposure to the safety CS led to a reduction in blood oxygenation level-dependent activity in the amygdala and an increased activity in the striatum (Pollak *et al*, 2010b). A potential direct regulatory effect of the dorsolateral prefrontal cortex on amygdala activity had been suggested based on diffusion-tensor imaging-based tractography analysis that indicated direct connections (Pollak *et al*, 2010b). Similarly, another imaging study showed that conditioned subjects, who had learned to associate one cue with a mild shock and a second cue (safety) with no shock, showed higher amygdala activation during the presentation of the aversive cue, whereas greater striatal activation was found in the presence of the safety cue. When the reinforcement contingencies were reversed, the neural activity pattern in response to the previous fear cue shifted from the amygdala to areas of the ventral prefrontal cortices and the striatum (Schiller *et al*, 2008). These results are consistent with specific and distinctive neural circuitries subserving learned safety, which involve at least the LA, the striatum, and regions of the PFC (Figure 4).

A necessary requirement to conclude that these structures are functionally connected to learned safety could come from site-specific lesions, which could be complemented by *in vivo* stimulation experiments demonstrating that the LA and/or CP are not only required but also sufficient to induce learned safety. Specific lesions in rats indicate that the central nucleus of the amygdala, which is critical for the acquisition and expression of initial fear-potentiated startle (FPS), is not necessary for the expression of conditioned inhibition (Falls and Davis, 1995; Jovanovic *et al*, 2012b; Kazama *et al*, 2012). In macaques, lesions of the amygdala in the neonatal brain has been shown to retard—but not completely abolish—the acquisition of learned fear and disrupt the processes of learned safety only in some of the adult animals (Falls and Davis, 1995; Jovanovic *et al*, 2012b; Kazama *et al*, 2012).

Other Brain Regions Potentially Involved in Learned Safety

The posterior insula, termed sensory insula (Si), exhibits convergent responses to simultaneous multisensory stimulation (Rodgers *et al*, 2008) and has afferent intracortical and thalamocortical as well as efferent amygdala connections (McDonald *et al*, 1999; Shi and Cassell, 1998). These characteristics led to the speculation that Si might be involved in safety learning. Indeed, inhibition evoked neuronal activity of the Si by muscimol, which blocked the safety effect only when applied during stressor exposure, but not when the Si was inhibited during later behavioral testing (Christianson *et al*, 2008; Christianson *et al*, 2011). However, although results of this study clearly suggest that Si is involved in the transmission or application of learned safety, Si cannot be concluded to be also involved in the acquisition of learned safety. To this end, the effects of Si lesions or functional inhibition of Si activity need to be determined in both pre- and postlearning paradigm. Thus, a confirmed causal involvement of Si for safety learning would be specifically interesting as lesions in rats suggested

that the insula is not necessary for fear learning (Romanski and LeDoux, 1992; Shi and Davis, 1999). This would also propose an exclusive involvement of this structure in learned safety, independently of learned fear.

Another brain region potentially involved in learned safety is the posterior intralaminar nucleus (PIN) of the thalamus. Post-training lesions of this structure have been shown to disrupt conditioned inhibition of learned fear (Waddell *et al*, 2003), implying that PIN is critically involved in the expression of learned safety. However, pretraining lesions of PIN are required to find out whether the PIN is also important for the acquisition of safety learning. After understanding that Si and PIN are important for safety learning, the following question arises: How could neural circuits in or between Si and/or PIN inhibit learned fear? From one side, Si receives multimodal sensory inputs and projects directly to the BLA complex (McDonald *et al*, 1999) and cortical sensory information is conveyed to the amygdala via insular cortex (Shi and Davis, 1999). On the other side, PIN along with the overlaying medial division of the medial geniculate, which is also the source of direct thalamo-LA projection of auditory fear CS information (Romanski and LeDoux, 1992), send monosynaptic projections to LA as well as to the portion of dorsal striatum (CP in rodents) that lies immediately dorsal to LA (Rogan and LeDoux, 1995). Thus, it can be speculated that the insula, Si, and PIN work in concert with BLA during safety learning, leading to the inhibition of BLA to the BNSTlv circuit, a critical brain region for learned fear (Davis, 1992; LeDoux, 2000; Shin *et al*, 2006) (Figure 4). With regard to the cortical control mechanisms, which appear as a prerequisite for the modulation of fear expression in learned safety, a first human imaging study suggests an involvement of the dorsolateral prefrontal cortex (Pollak *et al*, 2010b), which recently also has been shown to be activated during anticipation of the CS in a fear extinction paradigm (Kattoor *et al*, 2013).

The specific role of the ventromedial prefrontal cortex, central to the neural circuitry of fear extinction, in learned safety is still not sufficiently investigated. The infralimbic (IL) region of the medial prefrontal cortex (mPFC) has a role in the inhibition of inappropriate responding (Quirk *et al*, 2000); hence, it appears as a candidate brain region that is also involved in the neuronal circuitries mediating learned safety. Indeed, IL-lesioned animals have been shown to fail in the retardation test after training in a conditioned inhibition procedure, whereas IL lesions did not affect their summation test performance (Rhodes and Killcross, 2007). These data indicate that IL is not important for the acquisition of inhibitory associations between a stimulus and reward and the expression of the value of the conditioned inhibitor when placed in competition with an excitatory cue. However, a selective role for IL in the competition for behavioral control between the inhibitory and excitatory associations of a single stimulus is suggested. Moreover, considering the pivotal role of the mPFC for fear extinction and the fact that in animals only a single lesion study of the mPFC to examine a requirement of this important cortical control center for safety learning had been carried out (Gewirtz *et al*, 1997), it is too premature to exclude definitely an involvement of the mPFC in learned safety. Even more, the particular study that failed to reveal an effect of mPFC lesions on safety learning (Gewirtz *et al*, 1997) also reported no impact of mPFC lesioning on fear extinction, a finding that is in contrast to a host of other reports that do document such effects (Lebron *et al*, 2004; Morgan and LeDoux, 1995b; Morgan *et al*, 1993, 2003; Quirk *et al*, 2000; Rhodes and Killcross, 2004, 2007). As such, it can be speculated that the specific localization and/or extension of the lesions in the work of Gewirtz *et al* (1997) may account for the observed lack of an effect of mPFC impairment on safety learning. Moreover, considering the fact that several human imaging studies report activation of the mPFC in response to a safety signal (Dolan, 2007; Milad *et al*, 2006; Phelps and LeDoux, 2005) and that the mPFC is also recruited in several other paradigms involving emotional regulation (Etkin *et al*, 2006, 2011a; Etkin and Schatzberg, 2011b; Roy *et al*, 2012) invites a reassessment of the role of this prefrontal control center in learned safety.

Brain Regions Most Likely not Involved in Learned Safety

Apart from the involvement of LA, CP, and PFC in learned safety, the roles of other brain regions, including the nucleus accumbens (Nac) (Josselyn *et al*, 2005) and the perirhinal cortex (Prh) (Falls *et al*, 1997) have also been assessed.

It has been suggest that the Nac is not critically involved in learned safety or learned fear because neither the increasing dopaminergic nor the decreasing glutamatergic function in Nac altered learned safety or learned fear in rats (Josselyn *et al*, 2005). More critically, large pre- or post-training electrolytic lesions of the Nac did not affect the acquisition or expression of learned safety or learned fear (Josselyn *et al*, 2005). However, an indirect role for the Nac in fear extinction has recently been proposed based on the observation that deep-brain stimulation of the Nac rescued an impairment of deficient extinction retrieval in a genetic mouse model and this effect has been attributed to an interaction between the Nac and the corticolimbic extinction circuitry (Whittle *et al*, 2013). Similarly, deep brain stimulation of specific zone dorsomedial of the ventral striatum has been described to augment extinction of conditioned fear in rats (Rodriguez-Romaguera *et al*, 2012). It can be speculated that the involvement of the Nac specifically relates to fear extinction rather than learned safety, as its involvement appears to originate from an effect on particular prefrontal areas, which could be selectively recruited during fear extinction.

The Prh is located in a pivotal position to influence the flow of information into and out of the hippocampus, and hence, Prh is suggested to be associated with learned fear (Kealy and Commins, 2011; Milad *et al*, 2006; Rosen and Donley, 2006). Examining the function of the Prh in learned safety, a single study reports that—whereas post-training lesions suggest an involvement of Prh in learned fear—no evidence for effect of lesion on conditioned inhibition was obtained as Prh-lesioned animals retained the capability to inhibit the startle response induced by fear conditioning (Falls *et al*, 1997). However, a human imaging study suggests an involvement of the Prh in the neural circuitries of both fear extinction and latent inhibition and proposes that the involvement of the Prh may relate to its influence on the Nac (Puga *et al*, 2007).

MODULATORY INFLUENCES ON LEARNED SAFETY

In the case of mood and anxiety disorders that are of multifactorial etiology, only some but not all individuals exposed to chronic stress or psychological trauma develop the disease. Similarly, a variety of factors may contribute to the acquisition and expression of emotional responses in the corresponding animal paradigms. As is the case in the human population, where genetic factors account for much of the observed diversity, the genetic background of the model studied may be one of the major influencing factors on safety learning. Indeed, it has recently been shown that when 129SI/Svlmj (SI) and C57BL/6J (B6) inbred mice were compared, mice of the SI strain exhibited overgeneralized fear to conditioned stimuli and impaired ability to inhibit fear responses when the safety cue was presented (impaired safety learning), whereas the B6 strain behaved normally (Ostroff *et al*, 2010). These results provide evidence of a genetic contribution to the ability to identify and properly respond to environmental cues, predicting protection from danger.

In humans, twin studies have been used to estimate the influence of genetics in predisposed individuals on stress-related mood and anxiety disorders. Such studies have shown that genetic risk factors increased the probability for major depression and had a higher impact on females than males (Kendler, 2001). Furthermore, specific gene variants (eg serotonin transporter, pituitary adenylate cyclase-activating polypeptide (PACAP), and PAC1 receptor) have also found to influence the susceptibility as well as resilience to develop mood disorders (Mahan and Ressler, 2012; Uher, 2008). However, whether and how these and other genetic risk factors together with other factors, such as age, sex, and environmental conditions, may also have a role in regulating learned safety remain to be elucidated in future studies.

THE CELLULAR AND MOLECULAR MECHANISM OF LEARNED SAFETY

Although the original concept of learned safety—as a special case of conditioned inhibition—dates back to Pavlov, the underlying neurobiological mechanisms at the systemic, cellular, and molecular level have only recently begun to be elucidated.

Morphological Correlates of Learned Safety

In a previous study, opposing neural responses were observed in the LA and CP of safety and fear conditioned mice (Rogan *et al*, 2005). To examine the cellular morphological consequences in response to learned safety/fear in the LA, serial electron microscopy has been used to reconstruct dendrites after either fear or safety conditioning (Ostroff *et al*, 2010). It was found that learned safety tended to result in not only smaller spines synapses but also smaller spine apparatus, a smooth endoplasmic reticulum, inside the spine synapses, whereas learned fear led to larger spine apparatus as well as larger spine synapses. The enlarged spine apparatus and spine synapse after fear conditioning indicated an enhancement of stable neural connectivity, whereas safety learning may weaken synapses without destabilizing connections because conditioned inhibition is less robust than fear conditioning (Ostroff *et al*, 2010; Rescorla, 1969; Rogan *et al*, 2005). Structural changes in the amygdala have also been reported as a result of fear extinction learning where behavioral training induced clustering of GABA-A receptors in the synaptic cleft, hereby favoring utmost inhibition (Chhatwal *et al*, 2006). Moreover, enhanced expression of cell adhesion molecules required for stabilization induced by extinction learning have been observed in the BLA (Markram *et al*, 2007), further supporting the amygdala as central structure involved in the synaptic remodeling during both extinction and safety learning.

The Amygdala Gene Expression Profile of Learned Safety

A major advancement in the elucidation of the molecular mechanism of learned safety has been achieved through a study examining the gene expression profile in BLA of learned safety-trained mice (Pollak *et al*, 2008). Interestingly, it was found that although learned safety served as a behavioral antidepressant in two animal tests for depression (ie the forced-swim test and the sucrose preference test) (Pollak *et al*, 2008; Sevelinges *et al*, 2011)—in a manner comparable to the effect achieved by treatment with pharmacological antidepressants (such as the selective serotonin reuptake inhibitor fluoxetine)—alternative molecular signal-transduction pathways seemed to be mediating its antidepressant activity. Although most frequently described antidepressant drugs act upon the monoaminergic systems, specifically the serotonergic system (such as fluoxetine), learned safety seems to involve modulation of dopaminergic and neuropeptidergic signaling, specifically dopamine type 2 receptors and substance P together with other molecules previously implicated in stress response and the pathogenesis of depression, including preproenekphalin1 and prodynorphin (Pollak *et al*, 2008). Future molecular studies may address molecular signaling, providing the basis for an involvement of proposed neurotransmitter systems and examining the upstream events responsible for initiating the observed gene expression changes.

The Role of Neurogenesis in Learned Safety

Hippocampal neurogenesis seems to be pivotal for learned safety as mice with abolished hippocampal neurogenesis displayed retardation in the acquisition of learned safety (Pollak *et al*, 2008). Moreover, the antidepressant effect of learned safety is hindered in x-irradiated mice in which the ability to develop newborn cells in the hippocampal dentate gyrus is ablated and learned safety itself leads to enhanced survival of newborn cells in the dentate gyrus. These data suggest an important role of hippocampal neurogenesis for learned safety. Evidence for an enhancement of hippocampal brain-derived neurotrophic factor (BDNF) in learned safety (Pollak *et al*, 2008) proposes that this molecular modification might contribute to the antidepressant-like phenotype of learned safety by providing enhanced neurotrophic support to newly generated cells in the hippocampus. Within the neural network engaged in safety learning, it can be proposed that, similarity to recent observations for fear extinction (Sotres-Bayon *et al*, 2012),

the hippocampus might be required for gating the modulatory influences of prefrontal regions, leading to an inhibition of the emotional response orchestrated by the amygdala during learned safety (Sotres-Bayon *et al*, 2012). The deficiency in safety learning observed in animals with ablated hippocampal neurogenesis (Pollak *et al*, 2008) may therefore be based on an impairment of hippocampal inhibition of the prefrontal cortex and relate to the morphological and volumetric hippocampal aberrations reported in patients suffering from both, PTSD and depression (Kitayama *et al*, 2005; Videbech and Ravnkilde, 2004).

Also, the fact that the recall of the safety signal can undergo contextual modulation as it exerts its behavioral effects also in a novel context distinct of the original conditioning environment implies that hippocampal plasticity might be the neural requirement for this particular phase of safety learning. As is the case for other types of learning, also safety learning, from acquisition over consolidation to recall most likely involves various brain structures. The amygdala most likely constitutes the major site for acquisition and inhibition of the fear response, whereas the hippocampus may be necessary for the adjustment of the behavioral consequences of learned safety in an independent context. As such, the observed increased levels of hippocampal BDNF and augmented hippocampal neurogenesis (Pollak *et al*, 2008) are very probable candidate mechanisms mediating this contextual flexibility at the molecular level. Interestingly, also during fear extinction, specifically the extinction of contextual fear, where the hippocampus is most importantly involved in the retrieval of extinction memory (see for a review Quirk and Mueller, 2008), both hippocampal neurogenesis (Cleva *et al*, 2011; Deng *et al*, 2009; Ko *et al*, 2009; Pan *et al*, 2012) and BDNF expression (Andero and Ressler, 2012; Heldt *et al*, 2007; Liu *et al*, 2004) are required for its behavioral effects. As for recall of fear extinction, a critical role for BDNF expression has also been demonstrated in the rat BLA, where fear extinction learning also induced upregulation of its mRNA (Chhatwal *et al*, 2006), an intriguing observation that has not yet been tested for learned safety.

RELEVANCE OF LEARNED SAFETY IN BASIC NEUROSCIENCE

Conditioned fear is one of the most widely used animal models for studying the neurobiological basis of fear and anxiety disorders (Davis and Shi, 1999; LeDoux, 2000; Phelps and LeDoux, 2005). Conditioned inhibition of fear (or learned safety) is a relatively unexplored behavior paradigm exhibiting two different aspects of fear regulation. First, it represents an important modulatory system that prevents exaggerated emotional responses that are disproportional to the inducing stimulus or inappropriate at the specific circumstance. Second, the identification of protection from danger: both aspects are critical for self-preservation and well-being (Pollak *et al*, 2010a). Although the behavioral outcome of learned fear in rodents, usually, involves the display of freezing behavior as natural self-defense response, learned safety represents a more active approach to favoring survival in dangerous situations (ie, the identification of safe environments). Hence, the paradigm of learned safety, which has been successfully established in laboratory animals (Pollak *et al*, 2010a), enables us to study the neurobiological basis of this adaptive learning process. This helps to understand how learned safety acts to inhibit the responses evoked by learned fear at systemic, cellular, and molecular levels (Rogan *et al*, 2005). Moreover, learned safety represents an important extension of animal models for investigating not only the control (inhibition) of emotions but also positively affected emotional states.

The ability of the safety signal to induce an antidepressant-like phenotype in behavioral paradigms independent of the original conditioning procedure implies that the effect of learned safety is not restricted to the reduction of conditioned fear specific to the original learning context. Supporting this idea is an early report on the transfer of conditioned inhibition across different aversive reinforcers in the rat (Nieto and Posadas-Andrews, 1984) as well as description of the impact of a learned safety signal on innate anxiety, exploratory activity in the open field and place preference (Rogan *et al*, 2005). Consequently, learned safety signals themselves can become positive reinforcers and exert anxiolytic properties (Rogan *et al*, 2005), which, in turn, enables individuals to learn and take advantage of sources of safety and security in the environment.

TRANSLATIONAL ASPECTS AND POTENTIAL APPLICATIONS IN CLINICAL SCIENCES

As is the case for experimental research with animal models, learned safety is only beginning to be explored in humans (Grillon and Ameli, 2001; Grillon *et al*, 1994b; Jovanovic *et al*, 2010, 2012b; Lissek *et al*, 2009; Pollak *et al*, 2010b; Schiller *et al*, 2008). Already in the 1990s, pioneering work by Christian Grillon and Michael Davis firstly described the translational potential of learned safety by examining the impact of safety signals on human anxiety and found that safety signals were able to reduce anticipatory anxiety as revealed by a FPS paradigm (Grillon *et al*, 1994b). This seminal study is of great importance considering the translational value of the FPS response, a most commonly used parameter for fear in human fear-conditioning paradigms. The startle reflex, a motor response elicited by the presentation of an unexpected auditory stimulus, can be induced in all mammals, enhanced by presentations of a fear CS (FPS) and its relatively simple neural circuitry is well understood. Davis and Grillon then went on to further show that FPS is altered in several psychopathologies related to aberrant anxiety states including panic disorder (Grillon *et al*, 1994a), PTSD (Morgan *et al*, 1995a, 1996). An important leap forward in the field of safety learning and its translational aspects was therefore the description of FPS as a measure of fear inhibition resulting from safety learning in a human conditional discrimination paradigm (Jovanovic *et al*, 2005). This study translates a discrimination procedure in rats based on earlier learning theory experiments (Rescorla, 1971; Wagner *et al*, 1968) to humans and sets the basis for examining the role and potential alterations of learned safety in patients with affective disorders using FPS as a tool for objective assessment of

the safety response. As such, there are currently two paradigms available, which allow direct translation of an animal protocol of learned safety to people: the conditional discrimination paradigm and the explicitly unpaired procedure (Figure 3). Indeed as proof of principle, subsequent studies provided evidence of dysfunctional safety learning and neural processing specifically in patients suffering from PTSD (Jovanovic *et al*, 2012a, b;Norrholm *et al*, 2013) and even proposed impaired safety learning as a biomarker for PTSD (Jovanovic *et al*, 2012b), allowing differentiation of acute stress disorder from chronic post-traumatic stress disorder (Jovanovic *et al*, 2013). Importantly, the ability to properly distinguish between signals indicating danger and those predicting safety may act as an intermediate phenotype for the basic research on pathomechanisms of PTSD, as it relates to both the neural circuitry involved in the disease and the symptomatology presented by patients suffering from the disorder (Jovanovic *et al*, 2012b). Also, impaired fear extinction has been related to the pathophysiology of PTSD (Orr and Roth, 2000; Peri *et al*, 2000). As such, both processes, the exaggerated and persistent fear responses to cues relating to the traumatic event—reflected in deficiency to acquire extinction learning—and the inability to reduce this fear response, despite the presence of signals indicating a safe environment—mirrored in an impairment of safety learning—appear to relate to the different clinical features forming part of the symptomatology associated with PTSD and listed as the diagnostic criteria in the DSM-IV TR: deficient fear extinction may relate to criterion B, intrusive recollection (specifically point 5: 'Physiologic reactivity upon exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event.').

By contrast, impaired safety learning presumably shares aspects of the behavior described in criterion D, hyperarousal (specifically point 4 'Hypervigilance' and point 5 'Exaggerated startle response'), as the inability to properly respond to safety cues can lead to hypervigilance (Jovanovic *et al*, 2009; Jovanovic *et al*, 2010) and learned safety acts to reduce auditory FPS responses (see, eg, Jovanovic *et al*, 2005; Jovanovic *et al*, 2013). The fact that fear extinction and safety learning seem to model different symptomatological aspects of PTSD can be explained by their distinct neurobiological features, including neural circuitry and molecular signaling involved and suggests the two paradigms as complementing animal models in PTSD basic and translational research.

Although the response to learned safety signals had been found not affected in individuals suffering from major depressive disorders (MDD) in one study (Jovanovic *et al*, 2012a), the fact that learned safety induces neural activity patterns opposite to those observed in MDD (Pollak *et al*, 2010b) and an antidepressant-like effect in mice (Pollak *et al*, 2008), invites a more thorough assessment of learned safety in the context of depression and its relevance as a potential behavioral antidepressant. Exploring the underlying neurobiological mechanisms might allow for the discovery of novel therapeutic approaches or identification of alternative drug targets combating depressive disorders.

As is the case for fear extinction (see for a review Holmes and Singewald, 2013), investigations on the neurobiological basis underlying individual differences in processing of learned safety (Hartley *et al*, 2011) appears as a promising approach for identifying personal risk factors potentially contributing to the development of psychiatric disorders and paving the way for the discovery of potential alternative treatment approaches.

In summary, the analysis of learned safety in humans and the examination of involved neural circuitries bear great potential as a tool for enhancing our understanding of aberrant neural processing in several psychiatric disorders (including anxiety disorders and depression). Moreover, the fact that learned safety can be induced in humans and experimental animals by comparable protocols that allow building a translational bridge between human and animal studies provides the opportunity to relate directly findings at the cellular and molecular levels obtained in experimental animals to the results from human studies investigating neural activity patterns and *vice versa*.

CONCLUDING REMARKS

The aim of this review article was to provide a comprehensive overview of the emerging research on learned safety from the molecular level in experimental animals, to its translational aspect to human studies, and to highlight its potential as animal model in neuropsychiatric research.

As we are in the initial stages of understanding the behavioral states encompassing learned safety, our insight into the underlying neurobiological pathways and the neural circuitry involved is still limited. By contrasting with the related but distinct paradigms of learned fear and fear extinction as laid out in here, the specific characteristics of learned safety and its relevance as translatable animal model for a defined endophenotype of PTSD become evident. Specifically, learned safety, respectively, deficiencies in learned safety—reflected in the incapability to take advantage of sources of security and protection offered in the environment—is proposed as animal model to study aspects of hyperarousal' and 'hypervigilance' related to the symptomatology of PTSD in a preclinical setting. An expanded and further in-depth analysis of the neural circuitry involved in learned safety (as proposed in the model in Figure 4) may provide further insight into the aberrant neural processes mediating this behavioral state and identify potential points of contact to ameliorate the associated symptoms.

Moreover, as opposed to fear extinction, the relevance of learned safety as animal model additionally expands to its potential role as behavioral antidepressant. Further investigations on these antidepressant effects of learned safety and its neurobiological underpinnings may enhance our understanding of the pathophysiology involved in depression and also offer alternative approaches for the identification of novel pharmacological targets aimed at combating some of the most debilitating mental diseases.

FUNDING AND DISCLOSURE

DDP is supported by Austrian Science Fund (FWF): stand-alone project P22424 and member of the research network SFB35. The authors declare no conflict of interest.

REFERENCES

Andero R, Ressler KJ (2012). Fear extinction and BDNF: translating animal models of PTSD to the clinic. *Genes Brain Behav* **11**: 503–512.

Baker AG (1974). Conditioned inhibition is not the symmetrical opposite of conditioned excitation: a test of the Rescorla–Wagner model. *Learn Motiv* **5**: 369–379.

Bang SJ, Allen TA, Jones LK, Boguszewski P, Brown TH (2008). Asymmetrical stimulus generalization following differential fear conditioning. *Neurobiol Learn Mem* **90**: 200–216.

Bouton ME, Westbrook RF, Corcoran KA, Maren S (2006). Contextual and temporal modulation of extinction: behavioral and biological mechanisms. *Biol Psychiatry* **60**: 352–360.

Cándido A, González F, Brugada ID (2004). Safety signals from avoidance learning but not from yoked classical conditioning training pass both summation and retardation tests for inhibition. *Behav Process* **66**: 153–160.

Chhatwal JP, Stanek-Rattiner L, Davis M, Ressler KJ (2006). Amygdala BDNF signaling is required for consolidation but not encoding of extinction. *Nat. Neurosci.* **9**: 870–872.

Choi JS, Lindquist DH, Brown TH (2001). Amygdala lesions block conditioned enhancement of the early component of the rat eyeblink reflex. *Behav Neurosci* **115**: 764–775.

Christianson JP, Benison AM, Jennings J, Sandsmark EK, Amat J, Kaufman RD *et al* (2008). The sensory insular cortex mediates the stress-buffering effects of safety signals but not behavioral control. *J Neurosci* **28**: 13703–13711.

Christianson JP, Jennings JH, Ragole T, Flyer JG, Benison AM, Barth DS *et al* (2011). Safety signals mitigate the consequences of uncontrollable stress via a circuit involving the sensory insular cortex and bed nucleus of the stria terminalis. *Biol Psychiatry* **70**: 458–464.

Cleva RM, Wischerath KC, Olive MF (2011). Extinction learning and adult neurogenesis. *Neuropsychopharmacology* **36**: 360–361.

Davis M (1992). The role of the amygdala in fear and anxiety. *Annu Rev Neurosci* **15**: 353–375.

Davis M, Shi C (1999). The extended amygdala: are the central nucleus of the amygdala and the bed nucleus of the stria terminalis differentially involved in fear versus anxiety? *Ann NY Acad Sci* **877**: 281–291.

Deng W, Saxe MD, Gallina IS, Gage FH (2009). Adult-born hippocampal dentate granule cells undergoing maturation modulate learning and memory in the brain. *J Neurosci* **29**: 13532–13542.

Dolan RJ (2007). The human amygdala and orbital prefrontal cortex in behavioural regulation. *Philos Trans R Soc Lond Ser B* **362**: 787–799.

Etkin A, Alarcon JM, Weisberg SP, Touzani K, Huang YY, Nordheim A *et al* (2006). A role in learning for SRF: deletion in the adult forebrain disrupts LTD and the formation of an immediate memory of a novel context. *Neuron* **50**: 127–143.

Etkin A, Egner T, Kalisch R (2011a). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci* **15**: 85–93.

Etkin A, Schatzberg AF (2011b). Common abnormalities and disorder-specific compensation during implicit regulation of emotional processing in generalized anxiety and major depressive disorders. *Am J Psychiatry* **168**: 968–978.

Falls WA, Bakken KT, Heldt SA (1997). Lesions of the perirhinal cortex interfere with conditioned excitation but not with conditioned inhibition of fear. *Behav Neurosci* **111**: 476–486.

Falls WA, Davis M (1995). Lesions of the central nucleus of the amygdala block conditioned excitation, but not conditioned inhibition of fear as measured with the fear-potentiated startle effect. *Behav Neurosci* **109**: 379–387.

Ganguly K, Kleinfeld D (2004). Goal-directed whisking increases phase-locking between vibrissa movement and electrical activity in primary sensory cortex in rat. *Proc Natl Acad Sci USA* **101**: 12348–12353.

Gewirtz JC, Falls WA, Davis M (1997). Normal conditioned inhibition and extinction of freezing and fear-potentiated startle following electrolytic lesions of medical prefrontal cortex in rats. *Behav Neurosci* **111**: 712–726.

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

Grillon C, Ameli R, Goddard A, Woods SW, Davis M (1994a). Baseline and fear-potentiated startle in panic disorder patients. *Biol Psychiatry* **35**: 431–439.

Grillon C, Falls WA, Ameli R, Davis M (1994b). Safety signals and human anxiety: a fear-potentiated startle study. *Anxiety* **1**: 13–21.

Hartley CA, Fischl B, Phelps EA (2011). Brain structure correlates of individual differences in the acquisition and inhibition of conditioned fear. *Cereb Cortex* **21**: 1954–1962.

Heldt SA, Stanek L, Chhatwal JP, Ressler KJ (2007). Hippocampus-specific deletion of BDNF in adult mice impairs spatial memory and extinction of aversive memories. *Mol Psychiatry* **12**: 656–670.

Holmes A, Singewald N (2013). Individual differences in recovery from traumatic fear. *Trends Neurosci* **36**: 23–31.

Ito W, Pan BX, Yang C, Thakur S, Morozov A (2009). Enhanced generalization of auditory conditioned fear in juvenile mice. *Learn Mem* **16**: 187–192.

Jha SK, Brennan FX, Pawlyk AC, Ross RJ, Morrison AR (2005). REM sleep: a sensitive index of fear conditioning in rats. *Eur J Neurosci* **21**: 1077–1080.

Josselyn SA, Falls WA, Gewirtz JC, Pistell P, Davis M (2005). The nucleus accumbens is not critically involved in mediating the effects of a safety signal on behavior. *Neuropsychopharmacology* **30**: 17–26.

Jovanovic T, Blanding NQ, Norrholm SD, Duncan E, Bradley B, Ressler KJ (2009). Childhood abuse is associated with increased startle reactivity in adulthood. *Depress Anxiety* **26**: 1018–1026.

Jovanovic T, Ely T, Fani N, Glover EM, Gutman D, Tone EB *et al* (2012a). Reduced neural activation during an inhibition task is associated with impaired fear inhibition in a traumatized civilian sample. *Cortex* **49**: 1884–1891.

Jovanovic T, Kazama A, Bachevalier J, Davis M (2012b). Impaired safety signal learning may be a biomarker of PTSD. *Neuropharmacology* **62**: 695–704.

Jovanovic T, Keyes M, Fiallos A, Myers KM, Davis M, Duncan EJ (2005). Fear potentiation and fear inhibition in a human fear-potentiated startle paradigm. *Biol Psychiatry* **57**: 1559–1564.

Jovanovic T, Norrholm SD, Blanding NQ, Davis M, Duncan E, Bradley B *et al* (2010). Impaired fear inhibition is a biomarker of PTSD but not depression. *Depress Anxiety* **27**: 244–251.

Jovanovic T, Sakoman AJ, Kozaric-Kovacic D, Mestrovic AH, Duncan EJ, Davis M *et al* (2013). Acute stress disorder versus chronic posttraumatic stress disorder: inhibition of fear as a function of time since trauma. *Depress Anxiety* **30**: 217–224.

Jungling K, Seidenbecher T, Sosulina L, Lesting J, Sangha S, Clark SD *et al* (2008). Neuropeptide S-mediated control of fear expression and extinction: role of intercalated GABAergic neurons in the amygdala. *Neuron* **59**: 298–310.

Kaplan GB, Moore KA (2011). The use of cognitive enhancers in animal models of fear extinction. *Pharmacol Biochem Behav* **99**: 217–228.

Kattoor J, Gizewski ER, Kotsis V, Benson S, Gramsch C, Theysohn N *et al* (2013). Fear conditioning in an abdominal pain model: neural responses during associative learning and extinction in healthy subjects. *PLoS One* **8**: e51149.

Kazama AM, Heuer E, Davis M, Bachevalier J (2012). Effects of neonatal amygdala lesions on fear learning, conditioned inhibition, and extinction in adult macaques. *Behav Neurosci* **126**: 392–403.

Kealy J, Commins S (2011). The rat perirhinal cortex: a review of anatomy, physiology, plasticity, and function. *Progr Neurobiol* **93**: 522–548.

Kendler KS (2001). Twin studies of psychiatric illness: an update. *Arch Gen Psychiatry* **58**: 1005–1014.

Kitayama N, Vaccarino V, Kutner M, Weiss P, Bremner JD (2005). Magnetic resonance imaging (MRI) measurement of hippocampal volume in posttraumatic stress disorder: a meta-analysis. *J Affect Disord* **88**: 79–86.

Ko HG, Jang DJ, Son J, Kwak C, Choi JH, Ji YH *et al* (2009). Effect of ablated hippocampal neurogenesis on the formation and extinction of contextual fear memory. *Mol Brain* **2**: 1.

Lattal KM, Maughan DK (2012). A parametric analysis of factors affecting acquisition and extinction of contextual fear in C57BL/6 and DBA/2 mice. *Behav Process* **90**: 49–57.

Laxmi TR, Stork O, Pape HC (2003). Generalisation of conditioned fear and its behavioural expression in mice. *Behav Brain Res* **145**: 89–98.

Lebron K, Milad MR, Quirk GJ (2004). Delayed recall of fear extinction in rats with lesions of ventral medial prefrontal cortex. *Learn Mem* **11**: 544–548.

LeDoux JE (2000). Emotion circuits in the brain. *Annu Rev Neurosci* **23**: 155–184.

Lissek S, Rabin SJ, McDowell DJ, Dvir S, Bradford DE, Geraci M *et al* (2009). Impaired discriminative fear-conditioning resulting from elevated fear responding to learned safety cues among individuals with panic disorder. *Behav Res Ther* **47**: 111–118.

Liu IY, Lyons WE, Mamounas LA, Thompson RF (2004). Brain-derived neurotrophic factor plays a critical role in contextual fear conditioning. *J Neurosci* **24**: 7958–7963.

Madan V, Brennan FX, Mann GL, Horbal AA, Dunn GA, Ross RJ *et al* (2008). Long-term effect of cued fear conditioning on REM sleep microarchitecture in rats. *Sleep* **31**: 497–503.

Mahan AL, Ressler KJ (2012). Fear conditioning, synaptic plasticity and the amygdala: implications for posttraumatic stress disorder. *Trends Neurosci* **35**: 24–35.

Maren S (2000). Auditory fear conditioning increases CS-elicited spike firing in lateral amygdala neurons even after extensive overtraining. *Eur J Neurosci* **12**: 4047–4054.

Maren S, Quirk GJ (2004). Neuronal signalling of fear memory. *Nat Rev Neurosci* **5**: 844–852.

Maren S, Yap SA, Goosens KA (2001). The amygdala is essential for the development of neuronal plasticity in the medial geniculate nucleus during auditory fear conditioning in rats. *J Neurosci* **21**: RC135.

Markram K, Gerardy-Schahn R, Sandi C (2007). Selective learning and memory impairments in mice deficient for polysialylated NCAM in adulthood. *Neuroscience* **144**: 788–796.

Masuda Y, Odashima J, Murai S, Saito H, Itoh M, Itoh T (1994). Radial arm maze behavior in mice when a return to the home cage serves as the reinforcer. *Physiol Behav* **56**: 785–788.

McDonald AJ, Shammah-Lagnado SJ, Shi C, Davis M (1999). Cortical afferents to the extended amygdala. *Ann NY Acad Sci* **877**: 309–338.

Milad MR, Rauch SL, Pitman RK, Quirk GJ (2006). Fear extinction in rats: implications for human brain imaging and anxiety disorders. *Biol Psychol* **73**: 61–71.

Morgan CA 3rd, Grillon C, Southwick SM, Davis M, Charney DS (1995a). Fear-potentiated startle in posttraumatic stress disorder. *Biol Psychiatry* **38**: 378–385.

Morgan CA 3rd, Grillon C, Southwick SM, Davis M, Charney DS (1996). Exaggerated acoustic startle reflex in Gulf War veterans with posttraumatic stress disorder. *Am J Psychiatry* **153**: 64–68.

Morgan MA, LeDoux JE (1995b). Differential contribution of dorsal and ventral medial prefrontal cortex to the acquisition and extinction of conditioned fear in rats. *Behav Neurosci* **109**: 681–688.

Morgan MA, Romanski LM, LeDoux JE (1993). Extinction of emotional learning: contribution of medial prefrontal cortex. *Neurosci Lett* **163**: 109–113.

Morgan MA, Schulkin J, LeDoux JE (2003). Ventral medial prefrontal cortex and emotional perseveration: the memory for prior extinction training. *Behav Brain Res* **146**: 121–130.

Myers KM, Davis M (2007). Mechanisms of fear extinction. *Mol Psychiatry* **12**: 120–150.

Nieto J, Posadas-Andrews A (1984). Effects of chlordiazepoxide on food anticipation, drinking and other behaviors in food-deprived and satiated rats. *Pharmacol Biochem Behav* **20**: 39–44.

Norrholm SD, Jovanovic T, Smith AK, Binder E, Klengel T, Conneely K *et al* (2013). Differential genetic and epigenetic regulation of catechol-*O*-methyltransferase is associated with impaired fear inhibition in posttraumatic stress disorder. *Front Behav Neurosci* **7**: 30.

Orr SP, Roth WT (2000). Psychophysiological assessment: clinical applications for PTSD. *J Affect Disord* **61**: 225–240.

Ostroff LE, Cain CK, Bedont J, Monfils MH, Ledoux JE (2010). Fear and safety learning differentially affect synapse size and dendritic translation in the lateral amygdala. *Proc Natl Acad Sci USA* **107**: 9418–9423.

Pan YW, Chan GC, Kuo CT, Storm DR, Xia Z (2012). Inhibition of adult neurogenesis by inducible and targeted deletion of ERK5 mitogen-activated protein kinase specifically in adult neurogenic regions impairs contextual fear extinction and remote fear memory. *J Neurosci* **32**: 6444–6455.

Pavlov IP (1927). *Conditioned Reflexes*. Dover: New York, NY, USA.

Peri T, Ben-Shakhar G, Orr SP, Shalev AY (2000). Psychophysiologic assessment of aversive conditioning in posttraumatic stress disorder. *Biol Psychiatry* **47**: 512–519.

Phelps EA, LeDoux JE (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* **48**: 175–187.

Pollak DD, Monje FJ, Lubec G (2010a). The learned safety paradigm as a mouse model for neuropsychiatric research. *Nat Protoc* **5**: 954–962.

Pollak DD, Monje FJ, Zuckerman L, Denny CA, Drew MR, Kandel ER (2008). An animal model of a behavioral intervention for depression. *Neuron* **60**: 149–161.

Pollak DD, Rogan MT, Egner T, Perez DL, Yanagihara TK, Hirsch J (2010b). A translational bridge between mouse and human models of learned safety. *Ann Med* **42**: 115–122.

Puga F, Barrett DW, Bastida CC, Gonzalez-Lima F (2007). Functional networks underlying latent inhibition learning in the mouse brain. *NeuroImage* **38**: 171–183.

Quirk GJ, Mueller D (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* **33**: 56–72.

Quirk GJ, Russo GK, Barron JL, Lebron K (2000). The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *J Neurosci* **20**: 6225–6231.

Rescorla RA (1969). Conditioned inhibition of fear resulting from negative CS–US contingencies. *J Comp Physiol Psychol* **67**: 504–509.

Rescorla RA (1971). Summation and retardation tests of latent inhibition. *J Comp Physiol Psychol* **75**: 77–81.

Rhodes SE, Killcross AS (2007). Lesions of rat infralimbic cortex result in disrupted retardation but normal summation test performance following training on a Pavlovian conditioned inhibition procedure. *Eur J Neurosci* **26**: 2654–2660.

Rhodes SE, Killcross S (2004). Lesions of rat infralimbic cortex enhance recovery and reinstatement of an appetitive Pavlovian response. *Learn Mem* **11**: 611–616.

Rodgers KM, Benison AM, Klein A, Barth DS (2008). Auditory, somatosensory, and multisensory insular cortex in the rat. *Cereb Cortex* **18**: 2941–2951.

Rodriguez-Romaguera J, Do Monte FH, Quirk GJ (2012). Deep brain stimulation of the ventral striatum enhances extinction of conditioned fear. *Proc Natl Acad Sci USA* **109**: 8764–8769.

Rogan MT, LeDoux JE (1995). LTP is accompanied by commensurate enhancement of auditory-evoked responses in a fear conditioning circuit. *Neuron* **15**: 127–136.

Rogan MT, Leon KS, Perez DL, Kandel ER (2005). Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. *Neuron* **46**: 309–320.

Rogan MT, Staubli UV, LeDoux JE (1997). Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**: 604–607.

Romanski LM, LeDoux JE (1992). Equipotentiality of thalamo-amygdala and thalamo-cortico-amygdala circuits in auditory fear conditioning. *J Neurosci* **12**: 4501–4509.

Rosen JB, Donley MP (2006). Animal studies of amygdala function in fear and uncertainty: relevance to human research. *Biol Psychol* **73**: 49–60.

Roy M, Shohamy D, Wager TD (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn Sci* **16**: 147–156.

Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA (2008). From fear to safety and back: reversal of fear in the human brain. *J Neurosci* **28**: 11517–11525.

Sevelinges Y, Mouly AM, Raineki C, Moriceau S, Forest C, Sullivan RM (2011). Adult depression-like behavior, amygdala and olfactory cortex functions are restored by odor previously paired with shock during infant's sensitive period attachment learning. *Dev Cogn Neurosci* **1**: 77–87.

Shi C, Davis M (1999). Pain pathways involved in fear conditioning measured with fear-potentiated startle: lesion studies. *J Neurosci* **19**: 420–430.

Shi CJ, Cassell MD (1998). Cortical, thalamic, and amygdaloid connections of the anterior and posterior insular cortices. *J Comp Neurol* **399**: 440–468.

Shin LM, Rauch SL, Pitman RK (2006). Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Ann NY Acad Sci* **1071**: 67–79.

Sotres-Bayon F, Sierra-Mercado D, Pardilla-Delgado E, Quirk GJ (2012). Gating of fear in prelimbic cortex by hippocampal and amygdala inputs. *Neuron* **76**: 804–812.

Uher R (2008). The implications of gene–environment interactions in depression: will cause inform cure? *Mol Psychiatry* **13**: 1070–1078.

Videbech P, Ravnkilde B (2004). Hippocampal volume and depression: a meta-analysis of MRI studies. *Am J Psychiatry* **161**: 1957–1966.

Waddell J, Heldt S, Falls WA (2003). Posttraining lesion of the superior colliculus interferes with feature-negative discrimination of fear-potentiated startle. *Behav Brain Res* **142**: 115–124.

Wagner AR, Logan FA, Haberlandt K, Price T (1968). Stimulus selection in animal discrimination learning. *J Exp Psychol* **76**: 171–180.

Whittle N, Schmuckermair C, Gunduz Cinar O, Hauschild M, Ferraguti F, Holmes A *et al* (2013). Deep brain stimulation, histone deacetylase inhibitors and glutamatergic drugs rescue resistance to fear extinction in a genetic mouse model. *Neuropharmacology* **64**: 414–423.

Williams DAO, Bruce J, LoLordo, Vincent M (1992). A reevaluation of Rescorla's early dictums about Pavlovian conditioned inhibition. *Psychol Bull* **111**: 275–290.