Inhibition of Fear by Learned Safety Signals: A Mini-Symposium Review

John P. Christianson, *University of Colorado*
Anushka B. P. Fernando, *University of Cambridge*
Andrew Kazama, *Emory University*
Tanja Jovanovic, *Emory University*
Linnaea E. Ostroff, *New York University*
Susan Sangha, *University of California San Francisco*

Journal Title: Journal of Neuroscience Nursing
Volume: Volume 32, Number 41
Publisher: Society for Neuroscience | 2012-10-10, Pages 14118-14124
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1523/JNEUROSCI.3340-12.2012
Permanent URL: https://pid.emory.edu/ark:/25593/s7sn8

Final published version: http://dx.doi.org/10.1523/JNEUROSCI.3340-12.2012

Published in final edited form as:
J Neurosci. 2012 October 10; 32(41): 14118–14124. doi:10.1523/JNEUROSCI.3340-12.2012.

Inhibition of Fear by Learned Safety Signals: minisymposium review

John P. Christianson[1], **Anushka B. P. Fernando**[2], **Andy M. Kazama**[3], **Tanja Jovanovic**[4], **Linnaea E. Ostroff**[5], and **Susan Sangha**[6]

John P. Christianson: John.christianson@colorado.edu

[1]Department of Psychology & Neuroscience, University of Colorado Boulder, CO 80309-0345

[2]Department of Experimental, Psychology University of Cambridge

[3]Department of Developmental Cognitive Neuroscience, Yerkes National Primate Research Center, Emory University

[4]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine

[5]Center for Neural Science, New York University

[6]Ernest Gallo Clinic and Research Center, University of California, San Francisco

Abstract

Safety signals are learned cues that predict the non-occurrence of an aversive event. As such, safety signals are potent inhibitors of fear and stress responses. Investigations of safety signal learning have increased over the last few years due in part to the finding that traumatized persons are unable to utilize safety cues to inhibit fear, making it a clinically relevant phenotype. The goal of this review is to present recent advances relating to the neural and behavioral mechanisms of safety learning and expression in rodents, non-human primates and humans.

INTRODUCTION

Despite numerous advances, the pathophysiology of posttraumatic stress disorder (PTSD) is not sufficiently understood and current treatments are not always therapeutic. A hallmark feature of PTSD is the heightened expression of fear or anxiety in environments where it is not appropriate. This symptom has been conceptualized as a generalization of the fear conditioned during the traumatic experience that becomes resistant to extinction (Rauch, Shin & Phelps, 2006). While this model has significant empirical support, emerging evidence suggests that in addition to extinction learning, safety learning is impaired (Jovanovic et al., 2012). As opposed to danger learning where a cue is paired with aversive stimulation, safety learning involves associating distinct environmental stimuli (safety signals) with the *non-occurrence* of aversive events. This is distinguishable from the phenomenon of fear extinction. During extinction, the danger cue is presented without aversive reinforcement and over time a new association forms that the cue predicts the absence of the aversive event. Thus the extinction learning introduces a new memory that competes with the original danger association, which results in the inhibition of fear (Bouton, 2004). In contrast, learned safety signals inhibit fear responses to cues that are normally still paired with an aversive event when the safety signal is not present. As such, safety signals are only learned when the subject expects danger but it does not occur. More fundamental to the clinical importance of safety learning, distinction between safe and

Correspondence to: John P. Christianson, John.christianson@colorado.edu.

dangerous circumstances is critical to survival in all animals and it is this process that is impaired in PTSD. Thus, identifying the mechanisms of safety learning represents a significant goal for basic neuroscience that should inform future prevention and treatment of PTSD and other anxiety disorders.

A "safety signal" is a specific class of conditioned inhibitor; as a result of Pavlovian conditioning it prevents or reduces the expression of fearful behaviors normally observed in the presence of an excitatory conditioned stimulus (CS) that had been paired with aversive unconditioned stimuli (US). Thus the first requirement for a safety signal is that it must come to inhibit the conditioned fear response as a result of *learning*—stimuli that interfere with conditioned responses without training are called external inhibitors. In fact, many of the stimuli commonly used in fear conditioning experiments may act as external inhibitors to fear responses such as startle and freezing (Myers & Davis 2004). For example, in one of our laboratories white noise was found to reduce behavioral freezing to a danger CS without prior conditioning (Christianson, unpublished observations). A stimulus intended for inhibitory conditioning should be excluded from a safety learning protocol if evidence for external inhibition is found.

Safety signals have been trained in numerous procedures (Figure 1). The critical procedure for observing fear inhibition by a safety signal is called a "summation test" and was first demonstrated by Hammond (1967). The example in Figure 1 is the result of **A+/B−** training in which **A** trials concluded with a mild footshock and **B** trials did not. The predicted result is that cue **A** becomes a danger CS and cue **B** becomes a safety signal. Summation tests involve presentations of the putative safety signal in compound with a conditioned excitor. After **A+/B−** training a significant reduction in behavioral freezing during **AB** is seen. Once learned, it is difficult to retrain a safety signal as a danger CS. Evidence for this phenomenon occurs after the safety signal is paired with the aversive stimulus (Rescorla, 1969). Upon later presentation of the safety signal less conditioned fear is observed compared to that equally conditioned to a novel CS. Rescorla described this as a "retardation" test because the observation is that new learning to the conditioned inhibitor appears to be delayed.

Pavlov (1927) and Konorski (1948) made early and significant experimental and theoretical contributions to conditioned inhibition. Later, a 1969(b) paper by Rescorla identified the boundary conditions for establishing a conditioned inhibitor that have shaped modern behavioral and neuroscience research. The authors point interested readers to the aforementioned seminal works and a volume by Miller & Spear (1985) that discuss the nuances of conditioned inhibition. This review aims to introduce a broad audience to recent advances concerning the conditions under which safety signals are learned and utilized in rodent, non-human primate and human subjects. Here we highlight the novel contributions of the investigators who participated in a minisymposium panel titled "Inhibition of fear by learned safety signals" at the annual meeting of the Society for Neuroscience, as well as noteworthy discoveries of others.

NEURAL CIRCUITRY OF SAFETY SIGNALS

Considering that information about safety is only relevant when one anticipates danger, knowledge about the circuitry of danger is a prerequisite to understanding safety. Fearful behaviors such as enhanced startle, behavioral freezing, social avoidance, autonomic arousal, etc., are the product of a well-understood neural circuit. When presented with an US, such as a footshock, thalamic neurons transmit sensory information regarding the US and co-occurring environmental stimuli to the lateral amygdala (LA) where plasticity occurs, thereby linking the US with environmental stimuli that now predict danger. A series

of excitatory relays from the lateral to the basal amygdala (BA) and then to the central nucleus of the amygdala (CeA) mediate the expression of fearful behaviors. This is an oversimplified schema (see Kim & Jung, 2006; Tronson et al., 2012 for excellent reviews), but it nevertheless captures a wealth of data. The majority of research focused on neural mechanisms of safety learning has centered on a simple hypothesis: safety signals must inhibit the output of the amygdala (for exceptions see Weirtelak, Maier & Maier, 1992 and Watkins et al., 1998). Amygdala neurons in the fear circuit are under the inhibitory control of local GABAergic interneurons (Ehrlich et al., 2009), the medial intercalated neurons (Amano et al., 2010), and the infralimbic ventromedial prefrontal cortex (vmPFC, Milad & Quirk, 2002). In addition to these known inhibitory circuits, the amygdala receives input from most cortical regions, hippocampus, striatum and brain stem nuclei (Sah et al., 2003) thus several putative pathways may contribute to safety signal processing.

Modern studies of safety learning utilizing lesion and pharmacological manipulations offer a litany of brain structures deemed insufficient to affect safety learning. A significant effort made by the laboratory of Michael Davis and his students led to reports of null effects in nucleus accumbens (Josselyn et al., 2005), CeA (Falls and Davis, 1995), vmPFC (Gewirtz et al., 1997; Christianson et al., 2008), hippocampus (Heldt, Coover & Falls, 2002) and perirhinal cortex (Falls, Bakken & Heldt, 1997). None of these manipulations were sufficient to impair safety signaling. However, lesions to the auditory thalamus and superior colliculus, sensory inputs to the LA, did disrupt conditioned inhibition of fear (Heldt & Falls, 2006; Waddell, Heldt & Falls, 2003).

The LA receives direct inputs from auditory areas that process stimuli commonly used to train fear and safety associations. As the animal learns to associate new meanings with these stimuli, changes in these inputs can be studied. Rogan and colleagues (2005) recorded synaptic responses to an auditory tone in the LA of awake, behaving mice. As had been seen previously, tone-evoked responses were enhanced when animals were fear conditioned to the tone (Rogan, Stäubli & LeDoux, 1997). Interestingly, these responses weakened when the tone was trained as a safety signal. These data indicate that a safety signal may not only inhibit the output of the amygdala complex, but also reduce a sensory cue's ability to excite the LA. This could explain the delay in acquiring a fear association to a previously established safety signal, as the relevant inputs must be increased from below baseline. Another comparison of LA synapses after fear or safety training found that synapse size increased with fear conditioning and decreased with safety conditioning (Ostroff et al., 2010). Interestingly, while there was additional evidence of changes in synapse morphology and density with fear conditioning (Ostroff et al., 2012), only reduced synapse size was seen with safety conditioning.

Simply reducing the potential of a stimulus to excite fear responses is not sufficient to establish a safety signal. It is therefore unlikely that weakening of LA synapses is the central mechanism of safety signal learning (although strengthening of LA synapses is undoubtedly a central mechanism of fear conditioning). In addition to recording in the LA, Rogan and colleagues (2005) also recorded tone-evoked synaptic responses in the striatum. Here, responses were enhanced with safety conditioning and weakened with fear conditioning. This indicates that the safety signal excites a region associated with approach and reward, providing it with an activating function that may be involved in switching behavior from defensive avoidance to approach when the safety signal is provided.

Safety signals possess rewarding qualities (discussed in detail below) and this led Pollak and colleagues (2008) to test whether safety training would act as a behavioral "anti-depressant." Indeed, safety training led to changes in forced swimming behaviors and anhedonia that occur with chronic fluoxetine treatment. As with chronic fluoxetine, safety training

increased neurogenesis in the dentate gyrus of the hippocampus and x-irradiation, a treatment that prevents neurogenesis, prevented the anti-depressant-like effects of safety training. Important to the discussion of safety signal mechanisms, Pollak's data imply that the hippocampus contributes to safety signal processing, however additional studies will be required to determine if its role is in learning or in regulation of emotion.

As safety signals are potent inhibitors of fear there exists a hypothesis that safety signals can mitigate the consequences of intermittent stressors (Weiss, 1971; Mineka et al., 1984). In fact, providing a safety signal during an intermittent shock stressor significantly attenuated 1) the behavioral freezing response during the stressor, 2) the induction of Fos protein in the LA, posterior BA and bed nucleus of the stria terminalis (Christianson et al., 2011), and 3) the long-term anxiogenic effect of the stressor (Christianson et al., 2008). Since lesions to the amygdala completely abolish the consequences of intermittent shock stress (Maier et al., 1993), cortical structures projecting to the amygdala nuclei were considered candidates for these safety signal effects. The posterior insular cortex (IC) receives multimodal somatosensory input (Nieuwenhuys, 2012), exhibits convergent responses to simultaneous multisensory stimulation (Rodgers et al., 2008), and projects to the amygdala (McDonald et al., 1999). These characteristics led Christianson and colleagues to hypothesize that the IC would be involved in learned safety. Indeed, both excitotoxic lesions (Christianson et al., 2008) and reversible pharmacological lesions (Christianson et al., 2011) completely abolished the stress-mitigating effect of the safety signal. This discovery added an important new candidate structure involved in emotional regulation. Importantly, the IC is a site of functional and structural abnormalities in anxiety and PTSD (Paulus & Stein, 2006; Hughes & Shin, 2011).

SAFETY SIGNAL LEARNING IN HUMAN AND NON-HUMAN PRIMATES

An advantage of focusing on fear inhibition by safety signals is that the known neurobiology of fear provides the necessary groundwork to understand fear-related mental disorders. Excessive fear and anxiety, along with an inability to overcome these emotions, are some of the defining characteristics of many anxiety disorders, such as phobias, panic disorder and PTSD. Over-generalization of trauma-related stimuli or situations (i.e. an impaired ability to discriminate between danger and safety cues (Jovanovic et al., 2009; 2010) can lead to hyper-vigilance and exaggerated physiological responses that are part of the PTSD clinical presentation. For example, combat veterans with PTSD may not be able to suppress fear in response to a previously learned fearful cue (e.g. helicopter sound), even when surrounded by many safety signals (e.g. a different time and place from the combat zone).

Laboratory paradigms that specifically test safety signal processing offer an objective assessment of the clinical impairment that does not rely solely on the patient's self-report of symptoms. Jovanovic and colleagues have translated a conditional discrimination procedure (**AX+/BX−**) based on a rodent model (See Figure 2) in which healthy individuals readily acquire excitatory associations with cue **A** when paired with **X,** whereas **B** paired with **X** becomes a safety signal (Jovanovic et al., 2005). Safety signal learning is tested by pairing **A** with **B** on summation test trials, which show a decreased fear response compared to **AX** (Figure 2). Jovanovic and colleagues have used this paradigm in PTSD patients from combat (2009) and civilian populations (2010) and both demonstrate impaired fear inhibition; Figure 2 shows data from both PTSD samples combined. Importantly, this phenomenon is unique to trauma-related disorders and is not seen in co-morbid mental illnesses such as depression (Jovanovic et al, 2010). In PTSD patients safety signal deficiencies may appear as early as 30 days and as late as 10+ years after trauma exposure, indicating that it is a persistent biomarker of psychopathology (T. Jovanovic et al., personal communication). The underlying neurobiological mechanisms for impaired safety signal processing in PTSD may

involve an interaction between hyper-activity of the amygdala and impaired top-down emotional control by the vmPFC (Rauch et al., 2006). Hypoactivity and structural differences in the vmPFC are consistently observed in PTSD populations (Etkin & Wager, 2007; Hughes & Shin, 2011; Corbo et al., 2005). Thus the vmPFC is hypothesized to be a site of pathophysiology in PTSD and may contribute to impairments in safety signal use.

As noted above, lesions to vmPFC in rat do not interfere with safety signals (Gerwitz et al., 1997; Christianson et al., 2008), weakening the vmPFC hypothesis. However, translation to the clinical model is problematic due to structural differences between the rodent and human frontal cortex (for comparison see Ongur & Price, 2000). In contrast to rodents, rhesus macaques share similar frontal cortices to humans, making them indispensable translational models. In the non-human primate version of the **AX+/BX−** fear-potentiated startle paradigm (Winslow, Noble, & Davis, 2007), monkeys first learn to discriminate conditioned cues predicting an aversive puff of air (**AX**) from conditioned cues signaling safety (**BX**), and are then presented with the **AB** summation test. Similar to healthy rodents and humans, monkeys discriminate between **A** and **B** cues and show an attenuated startle response when presented with the **AB** compound (Figure 2). After the summation tests animals receive extinction trials in which **A** is presented without the aversive air puff. This powerful paradigm allows safety signal learning and fear extinction to be studied within the same subject.

A developmental lesion study was conducted to examine the long-term effects of early damage to the amygdala, hippocampus, or the orbital frontal cortex (Brodmann areas 11 and 13). Surprisingly, safety signal learning and fear extinction were intact despite early damage to these structures (Kazama et al., 2012). Presently we speculate that compensatory mechanisms occurred across development, which have been reported in other tasks (Glavis-Bloom, Kazama, & Bachevalier, 2008). Additionally, although area 13 of the orbital frontal cortex has been shown to share strong connectivity with the amygdala (see Price, 2007 and Barbas, 2007 for review), human neuroimaging suggests that medial areas such as Brodmann area 14 may be more involved in flexibly modulating fear (Schuff et al., 2010; Milad et al., 2009). Additionally, neurophysiological data from non-human primate models suggest that dopaminergic striatal neurons may provide crucial prediction error feedback in the presence of safety signals (Matsumoto & Hikosaka, 2009). Thus safety signal processing likely involves multiple brain areas that have only begun to be examined. Much work remains but the non-human primate **AX+/BX−** paradigm will likely permit tremendous advances as a translational tool.

SAFETY SIGNALS AS A REINFORCER

Safety signals provide relief from fearful states and may reinforce safety-seeking behavior. The relief experienced during the presentation of a safety signal may also motivate avoidance behavior symptomatic of many anxiety disorders. This possibility is relevant to anxiety disorders such as obsessive-compulsive disorder. For example, patients with obsessional checking rituals have reported a feeling of relief following the completion of checking behaviors (Roper, 1973). The reinforcing properties of a conditioned inhibitor of fear can be understood conceptually when considering Konorski's (1967) proposal of two opposing motivational systems that reciprocally inhibit one another, an aversive system and an appetitive system. He and others (Dickinson and Pearce, 1977, Gray 1987) argued that inhibition of the aversive system by a safety signal should disinhibit the appetitive system and would therefore be functionally (or behaviorally) equivalent to a direct excitor of the appetitive system. Fernando and colleagues tested this hypothesis in rats by training a Pavlovian conditioned inhibitor, an auditory stimulus presented in the absence of a mild footshock in one group, and a Pavlovian conditioned appetitive excitor, an auditory stimulus

paired with a sucrose pellet in another group. Fernando and colleagues hypothesized that if relief functions as reward it should support the acquisition of a new behavioral response with subjects responding purely for the presentation of the inhibitor in the absence of any footshock (A. B. P. Fernando, personal communication). The appetitive stimulus preferentially supported a new instrumental response that was potentiated with D-amphetamine whereas the safety signal did not support a new instrumental response and therefore did not demonstrate reinforcing properties.

Safety signals can also be provided contingent upon performance of an escape or avoidance response (Dinsmoor, 2001; Soltysik & Zielinski, 1962; Weisman et al., 1966, Weisman & Litner, 1966, Weisman & Litner, 1969; 1971). Such stimuli become associated with the relief from shock and should reinforce the avoidance response. Rats were trained on an instrumental lever press avoidance task in order to see whether an instrumentally trained safety signal would reinforce an already acquired avoidance response (A. B. P. Fernando et al, personal communication). Akin to Rescorla (1969a), rats preferentially responded in a two-lever avoidance task on the lever that both prevented shock and produced the safety signal. Rats continued to demonstrate this preferential responding on the lever producing the safety signal in extinction. Both tests provide clear evidence that an instrumentally trained safety signal possesses reinforcing properties. Despite the demonstration of this study and others that safety signals possess reinforcing properties, there are clearly multiple lines of information processing associated with safety signals.

Safety signals alert the organism to when the environment is safe thus promoting behaviors leading to natural rewards, such as feeding and mating, whereas danger signals inhibit these behaviors. One example that highlights how these antagonistic motivational systems can have bi-directional effects on behavior can be seen in rats that suppress lever-pressing for food when a danger cue is presented but increase lever-pressing for food when a safety cue is presented (Walasek, Wesierska & Zielinski 1995), demonstrating that safety signals can promote food seeking behavior. Interestingly, even *Drosophila* display a similar antagonistic avoidance-approach behavior, showing conditioned avoidance of a danger odor and conditioned approach to the same odor cue if it signifies safety (Tanimoto, Heisenberg & Gerber, 2004).

Normally, obtaining natural rewards involves activating exploratory behavior and several safety-conditioning protocols have demonstrated increases in exploratory behavior. Safety cues can increase exploration in an environment that is normally anxiogenic to mice, demonstrating that safety cues can take on anxiolytic properties and can even be used to condition a place preference (Rogan et al., 2005; Pollak et al., 2008). Distinguishing between a safety signal's fear-inhibiting versus rewarding qualities is not simple and the behavioral phenomena suggest overlap in the neural circuitry. Sangha and colleagues (S. Sangha, J. Z. Chadick, & P. H. Janak, unpublished data) developed a safety signal training protocol that allows the parallel investigation of fear, safety and reward learning in rats with simultaneous single unit recordings in the basal amygdala (BA) of freely behaving rats. As the rats flexibly switched between freezing, inhibition of freezing and reward seeking several populations of neurons emerged. As expected, many neurons changed (i.e. either increased or decreased) their firing preferentially to either the danger cue or the reward cue. Importantly, a population of neurons showed selective responses to the safety cue. Finally, there was another population of neurons in the BA that showed a similar change in firing rate in response to the safety cue *and* reward cue, implying that there is an overlap of safety and reward encoding in the BA. These data are the first to correlate amygdala single unit activity with safety signals and behavior *in vivo*. In addition to Konorski's 1967 proposal of the safety signal disinhibiting the appetitive system through its inhibition of the aversive

system the observed overlap in neural encoding of safety and reward cues suggests that a safety signal may directly activate the appetitive system.

CONCLUSION

Conditioned inhibition of fear is not a new concept to neuroscience, however it is surprising that in 85 years since Pavlov's seminal work very little is known regarding the neural mechanisms underlying this phenomenon. This may come as a surprise to those familiar with fear conditioning and fear extinction for which tremendous advances, from the anatomical tracts to the molecular cascades, have been made. One reason for the disparity is that there is no standardized approach to train a safety signal and the ones typically employed are complex and do not lend themselves as easily to modern neuroscience approaches as Pavlovian fear conditioning.

As we prepared the minisymposium panel and this manuscript, our goal was to bring safety learning to a broad audience and identify the large gaps in our current understanding. Despite the gaps, however, it is possible to suggest a framework for continued investigation of safety learning. We have addressed three components of safety signal processing: learning, fear inhibition in a summation test, and rewarding characteristics. First, safety learning protocols require learning of a danger CS and then that a safety stimulus is associated with the non-occurrence of danger. Such discrimination might involve generation of an expectation error signal and subsequent updating of predictive value of the safety signal. Numerous structures including dorsal striatum, PFC, periaqueductal gray and amygdala have been implicated in expectation errors (Belova et al., 2007; McNally, Johansen & Blair, 2011; Schultz and Dickinson, 2000); therefore, these structures are potential contributors to safety learning. Second, tests for safety learning involve flexible behavioral responses that switch as danger and safety cues are presented to the subject. This process depends on recall of the learned cues; the putative anatomical loci for storage of the safety signal are unknown. Behavioral flexibility depends upon contributions from vmPFC, orbital frontal cortex, and the striatum (Wolfensteller & Ruge, 2012; Murray & Izquierdo, 2007). Again, interactions of these structures with the amygdala would be required to switch behavior from avoidance to approach as danger and safety signals appear in the environment. Third, safety signals possess rewarding properties. Like the fear circuitry, reward circuitry is well understood involving the ventral tegmental area to the nucleus accumbens dopaminergic circuit (Koob, 1992) and interactions between these structures and the amygdala are important for goal-directed behaviors (Schoenbaum, 1998; 2003). Thus, multiple circuits must interact with the amygdala to acquire, recall and utilize safety signals.

Future studies of safety signal processing must account for the different roles of the amygdala nuclei and cell populations known to be involved in both danger and safety learning (e.g. Christianson et al., 2011; Ehrlich et al., 2009; Ostroff et al., 2010). Fortunately, technologies including *in vivo* electrophysiology, optogenetics (e.g. Tye et al., 2011), designer receptors exclusively activated by designer drugs (DREADD; Dong et al., 2010), and genetic tools allow for neural observations and manipulations that are both anatomically and temporally precise—a requirement for dissecting safety from danger. Furthermore, Pollak and her colleagues established a safety learning protocol that is translatable between mouse models and humans (Pollak et al., 2010a) providing the field with yet another powerful translational research tool (Pollak et al., 2010b). Equipped with a modern neuroscience toolkit and an important clinical correlate we expect the next 85 years will welcome many exciting developments in our understanding of safety signals.

Acknowledgments

The authors wish to acknowledge funding in support of the reviewed work from the the National Institutes of Health (MH093412, MH070129, MH092576, MH47840 MH088985, MH58846, MH086947, MH083583), NARSAD, MRC Case Studentship, the State of California for Medical Research on Alcohol and Substance Abuse through the University of California at San Francisco to Dr. Patricia H. Janak, Yerkes base grant (RR-00165), and the National Center for Research Resources (P51RR165) currently supported by the Office of Research Infrastructure Programs/OD P51OD11132).

References

1. Amano T, Unal CT, Paré D. Synaptic correlates of fear extinction in the amygdala. Nat Neurosci. 2010; 13:489–94. [PubMed: 20208529]
2. Barbas H. Flow of information for emotions through temporal and orbitofrontal pathways. J Anat. 2007; 211:237–249. [PubMed: 17635630]
3. Belova MA, Paton JJ, Morrison SE, Salzman CD. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. Neuron. 2007; 55:970–984. [PubMed: 17880899]
4. Christianson JP, Benison AM, Jennings J, Sandsmark EK, Amat J, Kaufman RD, Baratta MV, Paul ED, Campeau S, Watkins LR, Barth DS, Maier SF. The sensory insular cortex mediates the stress-buffering effects of safety signals but not behavioral control. J Neurosci. 2008; 28:13703–11. [PubMed: 19074043]
5. Christianson JP, Jennings JH, Ragole T, Flyer JGN, Bension AM, Barth DS, Watkins LR, Maier SF. 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. 2011; 70:458–464. [PubMed: 21684526]
6. Corbo V, Clément M-H, Armony JL, Pruessner JC, Brunet A. Size versus shape differences: contrasting voxel-based and volumetric analyses of the anterior cingulate cortex in individuals with acute posttraumatic stress disorder. Biol Psychiatry. 2005; 58:119–124. [PubMed: 16038682]
7. Bouton ME. Context and behavioral processes in extinction. Learn Mem. 2004; 11:485–94. [PubMed: 15466298]
8. Dickinson A, Pearce JM. Inhibitory interactions between appetitive and aversive stimuli. Psychological Bulletin. 1977; 84:690–711.
9. Dinsmoor JA. Stimuli inevitably generated by behavior that avoids electric shock are inherently reinforcing. Journal of the Experimental Analysis of Behavior. 2001; 75:311–333. [PubMed: 11453621]
10. Dong S, Allen JA, Farrell M, Roth BL. A chemical-genetic approach for precise spatio-temporal control of cellular signaling. Mol Biosyst. 2010; 6:1376–80. [PubMed: 20532295]
11. Ehrlich I, Humeau Y, Grenier F, Ciocchi S, Herry C, Lüthi A. Amygdala inhibitory circuits and the control of fear memory. Neuron. 2009; 62:757–71. [PubMed: 19555645]
12. Etkin A, Wager T. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. Am J Psychiatry. 2007; 164:1476–1488. [PubMed: 17898336]
13. Falls WA, Bakken KT, Heldt SA. Lesions of the perirhinal cortex interfere with conditioned excitation but not with conditioned inhibition of fear. Behav Neurosci. 1997; 111:476–86. [PubMed: 9189262]
14. Falls WA, Davis M. 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. 1995; 109:379–87. [PubMed: 7662148]
15. Gewirtz JC, Falls WA, Davis M. Normal conditioned inhibition and extinction of freezing and fear-potentiated startle following electrolytic lesions of medical prefrontal cortex in rats. Behav Neurosci. 1997; 111:712–26. [PubMed: 9267649]
16. Glavis-Bloom, CK.; Kazama, AM.; Bachevalier, J. Neuroscience Meeting Planner. Vol. 2008. Washington, DC: Society for Neuroscience; 2008. Paradoxical functional facilitation in rhesus macaques with neonatal hippocampal lesions on two stimulus-reward association tasks. Online

17. Gray, JA. The psychology of fear and stress. 2. Cambridge: Cambridge University Press; 1987.
18. Hammond LJ. A traditional demonstration of the properties of Pavlovian inhibition using differential CER. Psychonomic Science. 1967; 9:65–66.
19. Heldt SA, Coover GD, Falls WA. Posttraining but not pretraining lesions of the hippocampus interfere with feature-negative discrimination of fear-potentiated startle. Hippocampus. 2002; 12:774–86. [PubMed: 12542229]
20. Heldt SA, Falls WA. Posttraining lesions of the auditory thalamus, but not cortex, disrupt the inhibition of fear conditioned to an auditory stimulus. Eur J Neurosci. 2006; 23:765–79. [PubMed: 16487157]
21. Hughes KC, Shin LM. Functional neuroimaging studies of post-traumatic stress disorder. Expert Rev Neurother. 2011; 11:275–85. [PubMed: 21306214]
22. Josselyn SA, Falls WA, Gewirtz JC, Pistell P, Davis M. The nucleus accumbens is not critically involved in mediating the effects of a safety signal on behavior. Neuropsychopharmacology. 2005; 30:17–26. [PubMed: 15257308]
23. Jovanovic T, Kazama AM, Bachevalier J, Davis M. Impaired safety signal learning may be a biomarker of PTSD. Neuropharmacology. 2012; 62:695–704. [PubMed: 21377482]
24. Jovanovic T, Keyes M, Fiallos A, Myers KM, Davis M, Duncan EJ. Fear potentiation and fear inhibition in a human fear-potentiated startle paradigm. Bioll Psychiatry. 2005; 57:1559–1564.
25. Jovanovic T, Norrholm SD, Fennell JE, Keyes M, Fiallos A, Myers KM, Davis M, Duncan EJ. Posttraumatic stress disorder may be associated with impaired fear inhibition: relation to symptom severity. Psychiatry Res. 2009; 167:151–60. [PubMed: 19345420]
26. Jovanovic T, Norrholm SD, Blanding NQ, Davis M, Duncan E, Bradley B, Ressler KJ. Impaired fear inhibition is a biomarker of PTSD but not depression. Depression and Anxiety. 2010; 27:244–251. [PubMed: 20143428]
27. Kazama AM, Heuer E, Davis M, Bachevalier J. Effects of neonatal amygdala lesions on fear learning, conditioned inhibition, and extinction in adult macaques. Behav Neurosci. 2012; 123:392–403. [PubMed: 22642884]
28. Kim JJ, Jung MW. Neural circuits and mechanisms involved in Pavlovian fear conditioning: a critical review. Neurosci Biobehav Rev. 2006; 30:188–202. [PubMed: 16120461]
29. Konorski, J. Conditioned reflexes and neuron organization. Cambridge: Cambridge University Press; 1948.
30. Konorski, J. Integrative Activity of the Brain. University of Chicago Press; Chicago: 1967.
31. Koob GF. Drugs of abuse: anatomy, pharmacology and function of reward pathways. Trends Pharmacol Sci. 1992; 13:177–84. [PubMed: 1604710]
32. Maier SF, Grahn RE, Kalman BA, Sutton LC, Wiertelak EP, Watkins LR. The role of the amygdala and dorsal raphe nucleus in mediating the behavioral consequences of inescapable shock. Behav Neurosci. 1993; 107:377–88. [PubMed: 8484901]
33. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature. 2009; 459:837–841. [PubMed: 19448610]
34. McDonald AJ, Shammah-Lagnado SJ, Shi C, Davis M. Cortical afferents to the extended amygdala. Ann N Y Acad Sci. 1999; 877:309–338. [PubMed: 10415657]
35. McNally GP, Johansen JP, Blair HT. Placing prediction into the fear circuit. Trends Neurosci. 2011; 34:283–92. [PubMed: 21549434]
36. Milad MR, Pitman RK, Ellis CB, Gold AL, Shin LM, Lasko NB, Zeidan MA, Handwerger K, Orr SP, Rauch SL. Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. Biol Psychiatry. 2009; 66:1075–1082. [PubMed: 19748076]
37. Milad MR, Quirk GJ. Neurons in medial prefrontal cortex signal memory for fear extinction. Nature. 2002; 420:70–4. [PubMed: 12422216]
38. Miller, RR.; Spear, NS., editors. Information Processing in Animals: Conditioned Inhibition. Hillsdale, N.J: Erlbaum; 1985.
39. Mineka S, Cook M, Miller S. Fear conditioned with escapable and inescapable shock: effects of a feedback stimulus. J Exp Psychol Anim Behav Process. 1984; 10:307–323.

40. Murray EA, Izquierdo A. Orbitofrontal cortex and amygdala contributions to affect and action in primates. Ann N Y Acad Sci. 2007; 1121:273–96. [PubMed: 17846154]
41. Myers KM, Davis M. AX+, BX− discrimination learning in the fear-potentiated startle paradigm: possible relevance to inhibitory fear learning in extinction. Learn Mem. 2004; 11:464–475. [PubMed: 15254216]
42. Nieuwenhuys R. The insular cortex: a review. Prog Brain Res. 2012; 195:123–63. [PubMed: 22230626]
43. Ongur D, Price JL. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cerebral Cortex Mar. 2000; 10:206–219.
44. Ostroff LE, Cain CK, Bedont J, Monfils MH, Ledoux JE. Fear and safety learning differentially affect synapse size and dendritic translation in the lateral amygdala. Proc Natl Acad Sci U S A. 2010; 107:9418–23. [PubMed: 20439732]
45. Ostroff LE, Cain CK, Jindal N, Dar N, Ledoux JE. Stability of presynaptic vesicle pools and changes in synapse morphology in the amygdala following fear learning in adult rats. J Comp Neurol. 2012; 520:295–314. [PubMed: 21674493]
46. Paulus MP, Stein MB. An insular view of anxiety. Biol Psychiatry. 2006; 60:383–7. [PubMed: 16780813]
47. Pavlov, IP. Conditioned reflexes. London: Oxford University Press; 1927.
48. Pollak DD, Monje FJ, Zuckerman L, Denny CA, Drew MR, Kandel ER. An animal model of a behavioral intervention for depression. Neuron. 2008; 60:149–161. [PubMed: 18940595]
49. Pollak DD, Monje FJ, Lubec G. The learned safety paradigm as a mouse model for neuropsychiatric research. Nat Protoc. 2010; 5:954–62. [PubMed: 20431541]
50. Pollak DD, Rogan MT, Egner T, Perez DL, Yanagihara TK, Hirsch J. A translational bridge between mouse and human models of learned safety. Ann Med. 2010; 42:115–22. [PubMed: 20121549]
51. Price JL. Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Ann N Y Acad Sci. 2007; 1121:54–71. [PubMed: 17698999]
52. Rauch SL, Shin LM, Phelps EA. Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research-past, present, and future. Biol Psychiatry. 2006; 60:376–382. [PubMed: 16919525]
53. Rescorla RA. Establishment of a positive reinforcer through contrast with shock. Journal of Comparative and Physiological Psychology. 1969a; 67:260–263. [PubMed: 5785614]
54. Rescorla RA. Pavlovian Conditioned Inhibition. Psychol Bull. 1969b; 72:77–94.
55. Rodgers KM, Benison AM, Klein A, Barth DS. Auditory, somatosensory, and multisensory insular cortex in the rat. Cereb Cortex. 2008; 18:2941–51. [PubMed: 18424777]
56. Rogan MT, Leon KS, Perez DL, Kandel ER. Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. Neuron. 2005; 46:309–320. [PubMed: 15848808]
57. Rogan MT, Stäubli UV, LeDoux JE. Fear conditioning induces associative long-term potentiation in the amygdala. Nature. 1997; 390:604–7. [PubMed: 9403688]
58. Roper G, Rachman S, Hodgson R. An experiment on obsessional checking. Behaviour Research and Therapy. 1973; 11:271–277. [PubMed: 4727291]
59. Sah P, Faber ES, Lopez De Armentia M, Power J. The amygdaloid complex: anatomy and physiology. Physiol Rev. 2003; 83:803–34. [PubMed: 12843409]
60. Sangha S, Chadick JZ, Janak PH. Amygdala neurons respond differentially to safety, fear and reward cues. submitted.
61. Schoenbaum G, Chiba AA, Gallagher M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat Neurosci. 1998; 1:155–159. [PubMed: 10195132]
62. Schoenbaum G, Setlow B, Saddoris MP, Gallagher M. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron. 2003; 39:855–867. [PubMed: 12948451]
63. Schuff N, Zhang Y, Zhan W, Lenoci M, Ching C, Boreta L, Mueller SG, Wang Z, Marmar CR, Weiner MW, Neylan TC. Patterns of altered cortical perfusion and diminished subcortical integrity

in posttraumatic stress disorder: an MRI study. Neuroimage. 2010; 54:S62–S68. [PubMed: 20483375]

64. Schultz W, Dickinson A. Neuronal coding of prediction errors. Annu Rev Neurosci. 2000; 23:473–500. [PubMed: 10845072]
65. Soltysik S, Zielinski K. Conditioned inhibition of the avoidance reflex. Acta biologiae experimentalis. 1962; 22:157–167. [PubMed: 13978142]
66. Tanimoto H, Heisenberg M, Gerber B. Event timing turns punishment to reward. Nature. 2004; 430:983. [PubMed: 15329711]
67. Tronson NC, Corcoran KA, Jovasevic V, Radulovic J. Fear conditioning and extinction: emotional states encoded by distinct signaling pathways. Trends Neurosci. 2012; 35:145–55. [PubMed: 22118930]
68. Tye KM, Prakash R, Kim SY, Fenno LE, Grosenick L, Zarabi H, Thompson KR, Gradinaru V, Ramakrishnan C, Deisseroth K. Amygdala circuitry mediating reversible and bidirectional control of anxiety. Nature. 2011; 471:358–62. [PubMed: 21389985]
69. Waddell J, Heldt S, Falls WA. Posttraining lesion of the superior colliculus interferes with feature-negative discrimination of fear-potentiated startle. Behav Brain Res. 2003; 142:115–24. [PubMed: 12798272]
70. Walasek G, Wesierska M, Zielinski K. Conditioning of fear and conditioning of safety in rats. Acta Neurobiol Exp. 1995; 55:121–132.
71. Watkins LR, Wiertelak EP, McGorry M, Martinez J, Schwartz B, Sisk D, Maier SF. Neurocircuitry of conditioned inhibition of analgesia: effects of amygdala, dorsal raphe, ventral medullary, and spinal cord lesions on antianalgesia in the rat. Behav Neurosci. 1998; 112:360–78. [PubMed: 9588483]
72. Wiertelak EP, Maier SF, Watkins LR. Cholecystokinin antianalgesia: safety cues abolish morphine analgesia. Science. 1992; 256:830–3. [PubMed: 1589765]
73. Weisman RG, Denny MR, Platt SA, Zerbolio DJ Jr. Facilitation of extinction by a stimulus associated with long nonshock confinement periods. J Comp Physiol Psychol. 1966; 62:26–30. [PubMed: 5968274]
74. Weisman RG, Litner JS. The course of Pavlovian excitation and inhibition of fear in rats. J Comp Physiol Psychol. 1969a; 69:667–672. [PubMed: 5359140]
75. Weisman RG, Litner JS. Positive conditioned reinforcement of Sidman avoidance behavior in rats. J Comp Physiol Psychol. 1969b; 68:597–603.
76. Weisman RG, Litner JS. Role of the intertrial interval in Pavlovian differential conditioning of fear in rats. J Comp Physiol Psychol. 1971; 74:211–218. [PubMed: 5541545]
77. Weiss JM. Effects of coping behavior with and without a feedback signal on stress pathology in rats. J Comp Physiol Psychol. 1971; 77:22–30. [PubMed: 5166077]
78. Winslow JT, Noble PL, Davis M. AX+/BX− discrimination learning in the fear-potentiated startle paradigm in monkeys. Learn Mem. 2008; 15:63–66. [PubMed: 18230674]
79. Wolfensteller U, Ruge H. Frontostriatal mechanisms in instruction-based learning as a hallmark of flexible goal-directed behavior. Front Psychol. 2012; 3:192. [PubMed: 22701445]

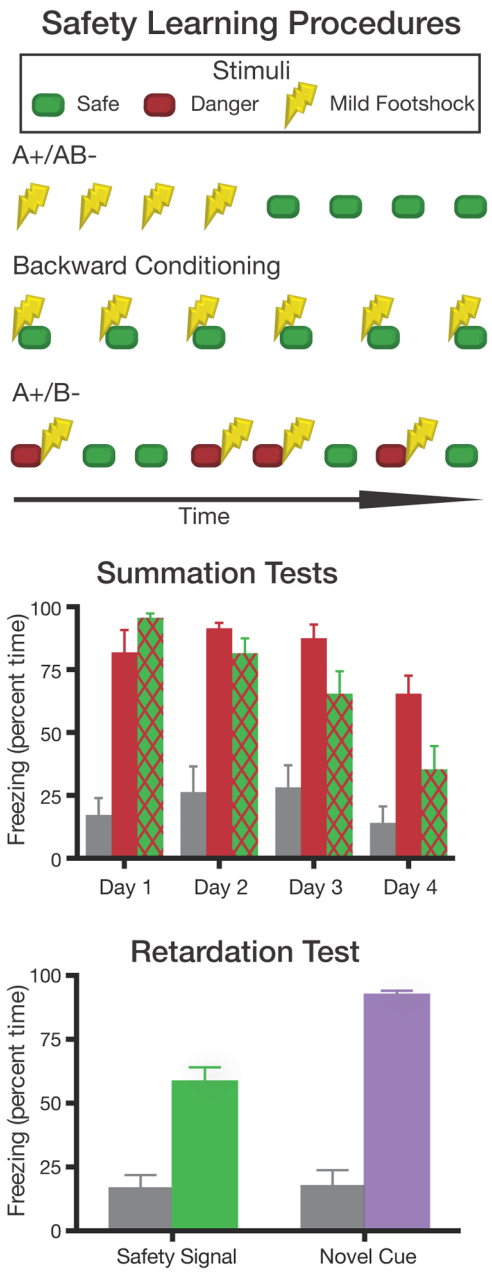

Figure 1. Examples of safety learning procedures and tests

Top: Among the many procedures used to condition a safety signal, most are related to the **A+/AB−**, backwards conditioned or **A+/B−** (discrimination training) designs. **A+/AB−** conditioning (see Rogan et al., 2005; Ostroff et al., 2010) involves separating the aversive cue from the putative safety signal in time by either clustering footshocks together (as shown) or interspersing safety signals between shocks with long inter-shock intervals. In a common experiment the **A** cue is the conditioning context and the **B** cue is a discrete stimulus (light or sound). In backwards conditioning the safety signal occurs just after termination of shock and so predicts the inter-shock period (Christianson et al., 2008). In **A+/B−** conditioning trials are presented in which a discrete cue **A** is always followed by shock and a discrete cue **B** is not. In **A+/B−** the context provides some expectation of shock during the **B** trials. Often a third cue (**X**) is added to both types of trials to transfer some expectation

of fear to the **B** trials (Myers & Davis, 2004). ***Middle***: Rats were trained on **A+/B−** (15 trials of each type) daily for 4 days. 24h after training rats returned to the training context and behavioral freezing was observed over 3 minutes. The first minute served as a baseline (grey bars), in the second minute the **A** cue was presented (red bars), and in the third minute the **B** cue was superimposed upon **A** (green & red hashed bars). Over days the **B** cue becomes a safety signal and inhibits freezing in the presence of **A**. ***Bottom:*** After 4 days of **A+/B−** training rats were then given 2 shock pairings of either the safety signal (**B**) or a novel cue. In a fear recall test 24h later (baseline freezing in grey bars), greater fear was observed to the novel cue (purple bar) than to the safety signal (green bar) indicating that safety training slowed new fear acquisition (example results from J. P. Christianson, unpublished data). These data are intended to provide an instructive example for readers unfamiliar with safety signal procedures and our laboratories have used all of these approaches to identify the neural mechanisms of safety learning.

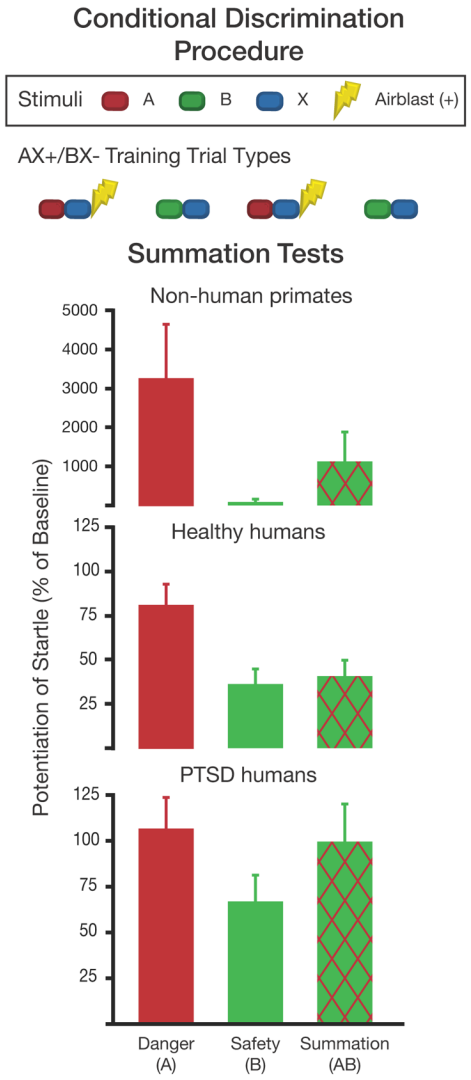

Figure 2. Conditional discrimination procedure with human and non-human primates

Top: The **AX+/BX−** conditioning phase includes presentations of **A** and **X** together predicting the US (airblast is used in the primate studies), while **B** and **X** together predict the non-occurrence of the US (Myers & Davis, 2004). ***Bottom:*** **A** and **B** are presented separately and then together during the post-training summation test; **B** acts as an inhibitor to reduce fear in the **AB** compound compared to **AX**. Intact rhesus macaques were trained on **AX+/BX−**, and show reduction in fear on the **AB** compound (Kazama et al, 2012). Similarly, healthy humans show a significant reduction of fear to **AB** relative to **A.** PTSD subjects discriminated between **A** and **B** but they did not inhibit startle during **AB** trials (data combined from Jovanovic et al, 2010, and T. Jovanovic et al, unpublished data).