Learning About Safety: Conditioned Inhibition as a Novel Approach to Fear Reduction Targeting the Developing Brain
Paola Odriozola, M.S., M.Phil., Dylan G. Gee, Ph.D.

Adolescence is a peak time for the onset of psychiatric disorders, with anxiety disorders being the most common and affecting as many as 30% of youths. A core feature of anxiety disorders is difficulty regulating fear, with evidence suggesting deficits in extinction learning and corresponding alterations in frontolimbic circuitry. Despite marked changes in this neural circuitry and extinction learning throughout development, interventions for anxious youths are largely based on principles of extinction learning studied in adulthood. Safety signal learning, based on conditioned inhibition of fear in the presence of a cue that indicates safety, has been shown to effectively reduce anxiety-like behavior in animal models and attenuate fear responses in healthy adults. Cross-species evidence suggests that safety signal learning involves connections between the ventral hippocampus and the prelimbic cortex in rodents or the dorsal anterior cingulate cortex in humans. Particularly because this pathway follows a different developmental trajectory than fronto-amygdala circuitry involved in traditional extinction learning, safety cues may provide a novel approach to reducing fear in youths. In this review, the authors leverage a translational framework to bring together findings from studies in animal models and humans and to bridge the gap between research on basic neuroscience and clinical treatment. The authors consider the potential application of safety signal learning for optimizing interventions for anxious youths by targeting the biological state of the developing brain. Based on the existing cross-species literature on safety signal learning, they propose that the judicious use of safety cues may be an effective and neurodevelopmentally optimized approach to enhancing treatment outcomes for youths with anxiety disorders.


Anxiety disorders, which are characterized by failure to regulate pathological fear, are the most common psychiatric illnesses, affecting up to one-third of the population (1). The majority of anxiety disorders are diagnosed in adolescence (2) and often persist into adulthood (3). Highlighting the need for intervention early in life, youths with anxiety disorders are at heightened risk for developing comorbid depressive and anxiety disorders, substance abuse, and suicidality (4–8). Although current evidence-based treatments can be highly effective for anxiety disorders (9, 10), up to 50% of both clinically anxious youths and adults do not respond sufficiently (i.e., they experience chronic anxiety or relapse following treatment) (8, 11–15). In this review, we consider the potential translation of safety signal learning for optimizing interventions for anxious youths by targeting the biological state of the developing brain. We first discuss principles of fear learning and evidence-based treatment for anxiety disorders, with a particular emphasis on youths. Next, we provide an overview of the literature on neural mechanisms of fear reduction in animal models and in humans, and describe how relevant neural circuitry changes with development and in anxiety disorders. Lastly, we introduce the concept of safety signal learning and discuss how translation across species and between basic science and practice can advance research and interventions for anxious youths.

FEAR LEARNING AND INTERVENTIONS FOR ANXIETY DISORDERS

Difficulty regulating fear is a central feature of anxiety disorders (16, 17). Fear learning is an adaptive function that allows an organism to predict potentially aversive events from environmental cues (see Table 1 for a glossary). When a stimulus no longer signals threat, fear expression gradually diminishes, allowing individuals to override the previously learned association (18). In classical conditioning, fear extinction (19) is operationalized by repeated exposure of the previously learned threat cue (i.e., conditioned stimulus, CS) without any aversive outcome (i.e., unconditioned stimulus, US) and is tested using threat and safety discrimination during extinction (i.e., extinction learning) or after a delay (i.e., extinction retention or recall). Deficits in extinction
TABLE 1. Glossary of terms related to safety signal learning

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fear learning</td>
<td>An adaptive function that allows an organism to predict potentially aversive events from environmental cues.</td>
</tr>
<tr>
<td>Unconditioned stimulus, US</td>
<td>A stimulus that reliably produces a natural, or unconditioned, response. In fear conditioning, this stimulus is aversive in nature (e.g., shock, aversive noise).</td>
</tr>
<tr>
<td>Conditioned stimulus, CS</td>
<td>A previously neutral stimulus that is paired with a US in classical conditioning (acquisition phase).</td>
</tr>
<tr>
<td>Fear extinction</td>
<td>The gradual process of fear reduction in classical conditioning that involves repeated exposure to the CS without the US.</td>
</tr>
<tr>
<td>Safety learning</td>
<td>The process of learning about safety in the environment. Tested using threat and safety discrimination.</td>
</tr>
<tr>
<td>Safety signal learning</td>
<td>The process through which a stimulus that is overly trained to signal the absence of threat (i.e., the safety cue) reduces fear in the presence of a threatening cue. Tested using the summation and retardation tests.</td>
</tr>
<tr>
<td>Conditioned inhibitor</td>
<td>A stimulus that inhibits a response as a result of learning (i.e., acquisition).</td>
</tr>
<tr>
<td>External inhibitor</td>
<td>A stimulus that inhibits a response without having undergone explicit training.</td>
</tr>
<tr>
<td>Safety behavior</td>
<td>An action or behavior that involves the use of a safety cue.</td>
</tr>
<tr>
<td>Preventive safety behavior</td>
<td>An action that employs a safety behavior prior to an anticipated threatening event to avoid harm or reduce the intensity of the event.</td>
</tr>
<tr>
<td>Restorative safety behavior</td>
<td>An action that employs a safety behavior following an aversive event with the goal of restoring safety.</td>
</tr>
</tbody>
</table>

Learning or extinction retention can result in unremitting fear that persists even after threat has subsided (20).

The primary evidence-based psychosocial treatment for anxiety disorders is exposure-based cognitive-behavioral therapy (CBT), which was initially based on models such as corrective information processing (21, 22) and is thought to rely on principles of fear extinction (22, 23). During exposures, patients repeatedly and systematically confront fear-provoking stimuli with the goal of reducing anxiety (9, 24). Such extinction-based interventions may be especially vulnerable to the return of extinguished fear responses, as the formation of a new extinction memory does not overwrite the initial fear association. Rather, the extinction memory competes with the original threat memory, and previously extinguished fear responses can return via the mere passage of time (spontaneous recovery), exposure to a stressor (fear reinstatement), or return to a fear-associated context (fear renewal) (25).

The challenges of sustained efficacy with CBT may be further compounded for youths with anxiety disorders. A high proportion of clinically anxious youths do not remit after gold-standard treatments (i.e., CBT and selective serotonin reuptake inhibitors [SSRIs]), with estimates ranging from 25% to 50% of youths still meeting criteria for a principal anxiety disorder after treatment (11, 26). A large-scale study of pediatric anxiety (12) found that only approximately 22% of youths achieved stable remission (i.e., did not meet diagnostic criteria for any anxiety disorder) across 4 years of annual follow-up that began 4–12 years after initial randomization to one of four 12-week treatment options (CBT, an SSRI, combined CBT and SSRI, or pill placebo). Within the group randomized to CBT alone, remission rates ranged from 40% to 60% in each follow-up year, highlighting the need to optimize treatments for pediatric anxiety.

Although response rates for CBT are similar across childhood, adolescence, and adulthood (11, 12, 14, 26), the factors contributing to low response rates may differ across age groups. In youths, one hypothesis is that some children and adolescents do not benefit sufficiently from current treatments because these interventions are largely based on principles that have been studied and implemented in adults. Delineating the biological state of the developing brain—that is, the nature of neural structure, function, and connectivity at a given developmental stage—and applying this knowledge to intervention approaches may be critical to optimizing treatment for anxious youths (27). Given evidence for diminished fear extinction learning during adolescence (28, 29), as well as altered ventromedial prefrontal cortex (vmPFC) and amygdala involvement during extinction learning in healthy and anxious adolescents (30), adolescents with anxiety may particularly benefit from efforts to optimize exposure-based therapies with approaches to fear reduction that either augment the strength of fronto-amygdala connections or bypass prefrontally mediated extinction processes.

NEURAL MECHANISMS OF FEAR REDUCTION

Identifying the neural mechanisms underlying core features of anxiety disorders is essential for the discovery of novel therapeutics and optimized treatments. Extensive research across species has shown that cortical-subcortical interactions primarily involving the amygdala, vmPFC, and hippocampus are central to fear learning and extinction learning and retention (23, 31–34). The amygdala is critically involved in fear learning and fear expression (35, 36), as well as extinction learning (37). Bidirectional connections between the amygdala and vmPFC modulate fear expression (38). Whereas the prelimbic (PL) region of the rodent vmPFC is involved in fear maintenance (39), the infralimbic (IL) region of the rodent vmPFC inhibits fear expression and stores and retrieves extinction memory (40). Consistent with these
findings, the IL cortex has been implicated in both extinction learning and retention (for reviews, see 38, 41, 42). The hippocampus has also been highlighted as a central region involved in fear learning and extinction and supplies critical information about the degree of threat or safety in the environment (43–45) via projections from ventral CA1 hippocampus to the basolateral amygdala (BLA) (46, 47). Additionally, recent molecular evidence may link learning about safety in the environment (i.e., safety learning) and extinction learning, including evidence that administering a cannabinoid receptor 1 antagonist in the hippocampus in adult rodents prevents extinction of avoidance behavior via safety learning (48). Furthermore, recent evidence suggests that glucocorticoid regulation enhances fear extinction learning and safety learning in individuals with posttraumatic stress disorder (PTSD) (49), further linking these processes at the molecular level with the endocannabinoid system and glucocorticoid hormones.

Neuroimaging studies are providing increasing insight into the neural circuitry supporting fear learning and extinction learning and recall in humans. Consistent with findings in rodents, evidence suggests that the amygdala plays a central role in fear learning in humans (23, 50–52) and that the dorsal anterior cingulate cortex (dACC) may modulate fear expression (53). Also in line with findings in rodents, human neuroimaging studies indicate that the amygdala also contributes to extinction learning (23, 52), whereas the vmPFC and hippocampus are involved in extinction retention (23, 54). Importantly, human neuroimaging studies must be considered in the context of the challenges and limitations that the field currently faces. Low reproducibility, low statistical power (e.g., stemming from a large number of dependent variables but a relatively small number of observations [subjects]), small effect sizes, and flexibility in data analyses have all been noted as limitations on the conclusions that can be drawn from neuroimaging studies (55–57). Thus, inferences about the neural circuitry supporting fear reduction in humans that are drawn from neuroimaging studies should be interpreted with caution, and it will be important that future studies employ evolving guidelines for best practices for reproducible science (57, 58). While studies are needed to replicate human neuroimaging findings with larger samples and greater statistical power, the existing evidence suggests involvement of frontolimbic circuitry in fear learning and extinction learning and retention, consistent with evidence in rodents.

Disruption in fear learning and extinction is a key etiological feature of anxiety disorders (59). Across studies, adults with anxiety disorders show increased subjective anxiety, skin conductance response, and startle response to safety cues during fear acquisition and increased fear responding to threat cues during extinction learning relative to individuals without anxiety (60, 61). Furthermore, both normative and clinically impairing anxiety involve alterations in the frontolimbic circuitry that underlies fear extinction (62). In human neuroimaging studies, adults with anxiety disorders display diminished prefrontal control of the amygdala, amygdala hyperreactivity, and weaker connectivity between the amygdala and various regions of prefrontal cortex, along with some evidence of altered hippocampal activation in studies of affective stimuli (62–68). Consistent with these disruptions, evidence has demonstrated alterations in this circuitry during extinction retention in anxiety disorders. Findings suggest impaired extinction retention (69) and functional alterations in the vmPFC, hippocampus, and dACC during extinction recall (63) among individuals with anxiety disorders.

Fear learning and extinction undergo dynamic changes during childhood and adolescence (29, 70; for a review, see 71). Neuroimaging studies in humans have begun to delineate age-related patterns of functional activation and connectivity during fear learning (72–74), extinction learning (30), and extinction recall (75–77). These findings suggest that changes in frontolimbic circuitry with age may contribute to changes in fear learning and extinction across development (78, 79). Cross-species evidence has suggested that fear extinction learning is diminished during adolescence (28, 29), relative to children and adults. This reduced extinction learning during adolescence has been associated with altered neuroplasticity in the vmPFC, namely, an absence of extinction-learning-induced plasticity within the rodent IL cortex (29). Furthermore, a recent study in healthy adolescents and young adults showed altered age-related involvement of the amygdala and vmPFC during extinction learning, such that younger adolescents showed higher amygdala activity and later engagement of the vmPFC to threat versus safety cues during extinction learning (30). These findings may be consistent with broader evidence of age-related changes in fronto-amygdala circuitry during childhood and adolescence (e.g., 80–85).

Although less is known about anxiety-related alterations to developmental trajectories, evidence suggests that youths with anxiety disorders may exhibit alterations in fear and extinction learning (86). Relative to nonanxious youths, anxious youths display increased self-reported fear (87) and skin conductance (88, 89) to both threat and safety cues and are more resistant to fear extinction, measured using startle response (90) and skin conductance (89). However, consistent evidence that anxious youths discriminate between threat and safety differently than nonanxious youths during conditioning is lacking (87, 91; for a review, see 92). For example, a recent meta-analysis of seven fear conditioning studies found that anxious youths exhibit stronger fear responses (i.e., self-reported fear, skin conductance response, or fear-potentiated startle) to individual threat and safety cues than nonanxious youths, but that differential fear acquisition and extinction (i.e., responding to threat versus safety cues) are similar between anxious and nonanxious youths (93). Given the small number of studies and participants in studies of fear and extinction learning in anxious youths, these results should be interpreted with caution, and further research is needed to understand these processes in pediatric anxiety disorders.
Building on behavioral and physiological studies in youths with anxiety, a small but growing body of research has investigated neurobiological alterations during fear learning and extinction learning and retention in anxious youths. Relatively few of these studies have conducted fear conditioning during fMRI scanning. However, initial results suggest that anxious adolescents show lower activation in the medial prefrontal cortex (PFC)/paracingulate gyrus in response to the safety cue (72). In addition to this alteration that was consistent across age, relative to nonanxious youths, anxious youths exhibited a stronger pattern of age-related decrease in prefrontal activation to the safety cue during fear conditioning (72). Neuroimaging studies have provided growing insight into neural processes related to retention of fear extinction in anxious youths. Anxious adolescents show lower subgenual anterior cingulate cortex activation, relative to healthy counterparts, when asked to indicate whether or not they were afraid of the stimuli (i.e., threat appraisal) during extinction recall (75). Furthermore, adolescents with anxiety also show higher vmPFC activation to the most prototypical threat and safety cues along a continuum, compared with anxious adults and nonanxious youths, when engaging in threat appraisal during extinction recall (75). One additional study using the same paradigm found that anxious youths showed higher negative amygdala-PFC functional connectivity when engaged in both threat appraisal and explicit threat memory during extinction recall, relative to anxious adults and nonanxious youths (76). Building on this work, a recent study from the same group (77) used the same extinction recall paradigm but increased the number of stimulus replicates and sample size (N=200) to increase statistical power. Findings showed that differences between anxious and nonanxious individuals were dependent on age. Specifically, anxious individuals showed an age-related decrease in amygdala-vmPFC functional connectivity compared with nonanxious individuals during extinction recall as stimuli increasingly resembled safety cues. The directionality of these results differed from the previous finding of higher amygdala-vmPFC connectivity in anxious youths (76); however, both findings highlight the relevance of this circuitry for extinction recall in anxious adolescents. While further human neuroimaging studies (e.g., with larger sample sizes) are needed to replicate these findings in anxious adolescents and to reconcile inconsistent findings, these studies suggest that youths with anxiety exhibit alterations in neural regions involved in fear learning and extinction recall and that these anxiety-related alterations may be developmental in nature.

**HARNESSING TRANSLATIONAL RESEARCH TO OPTIMIZE TREATMENTS**

Several approaches to enhancing fear reduction have been proposed, including reconsolidation update, in which the fear memory is recalled prior to extinction with the aim of updating the original fear memory trace (94, 95); combined cue and contextual fear extinction, in which extinction occurs in the original conditioning context (96); and safety signal learning, which is the process through which an organism learns about safety (or the lack of threat) in the environment (97). In this review, we focus on safety signal learning because of its potential capacity for translation that aims to optimize exposure-based therapies for fear reduction in a neurodevelopmentally informed manner. We aim to bridge the gap between cross-species research on basic neuroscience and its application to clinical treatment. The material that follows below comprises three sections: first, we review the literature on safety signal learning through conditioned inhibition and its potential neural correlates in rodents, nonhuman primates, and humans; next, we critically examine the current empirical literature on safety cues in treatment for anxiety, including the mechanisms and conditions under which safety cues may enhance or interfere with treatment; and, finally, we propose the application of safety signal learning to inform clinical translation that will augment evidence-based treatments for anxious youths.

Our proposed model highlights two particular aspects of translational research—translation between animal models and human research, and between basic neuroscience and clinical practice. Both of these levels of translation have critical importance for guiding future research and the treatment of anxiety disorders. For example, human neuroimaging findings in anxious populations can lead to circuitry-focused work in animal models; reciprocally, circuitry-focused work in experimental organisms is essential to delineating mechanisms of anxiety and can guide the selection of regions of interest in human neuroimaging research. Furthermore, neuroscientific findings across species can identify treatment targets or provide key insights that inform novel therapeutic techniques. Methodical implementation of interventions with neuroimaging can be used to evaluate the effectiveness of treatment and the utility of biomarkers. Taken together, findings from the basic science of safety signal learning across species and from treatment studies can reciprocally inform one another to guide the optimization of interventions for anxious youths by targeting the biological state of the developing brain. Throughout the review, we propose hypotheses based on the existing cross-species and clinical literature and provide recommendations for future translational research on safety signal learning.

**SAFETY SIGNAL LEARNING AS AN APPROACH FOR FEAR REDUCTION**

During safety signal learning, a cue that is overly trained to signal the absence of threat (i.e., the safety cue) is used to reduce fear in the presence of a threatening cue (97–99). Safety signal learning is a special class of conditioned inhibition in which the safety cue must inhibit the conditioned response (CR) as a result of learning (as opposed to the process by which stimuli inhibit the CR without training, called external inhibition) (97). Conditioned inhibition via
safety signal learning is thereby a process in which a stimulus is trained, through Pavlovian conditioning, to signal safety (or the absence of threat), and as a result this safety cue can inhibit the conditioned fear response. During the acquisition phase of conditioned inhibition, the threat cue (CS+) is paired with the US, and the safety cue (CS−) is never paired with the US. Research in rodents and nonhuman primates has relied on two procedures to test whether a stimulus acts as a conditioned inhibitor. During the “summation test,” the threat and safety cues are presented simultaneously as a compound stimulus (safety compound), yielding a reduction in fear-related behavior. During the “retardation test,” the safety cue is paired with the US (100). If a safety cue has been effectively learned, fear responding should be slower to emerge (relative to initial conditioning).

Safety signal learning via conditioned inhibition differs in several important ways from extinction learning or safety discrimination, which have been traditionally examined in relation to fear learning and anxiety disorders. Safety cues in Pavlovian conditioning are typically operationalized using a CS− that predicts the absence of a threatening stimulus (101). Safety discrimination can be measured during extinction learning, extinction recall, or reversal by comparing reactivity to the CS+ (threat cue) and the CS− (safety cue) (101). A large body of research in rodents (102) and humans (16, 103, 104) has investigated discrimination between threat and safety cues in extinction recall during development. However, these safety cues were not trained as conditioned inhibitors (i.e., the safety and threat cues were not paired or evaluated with a summation or retardation test). In contrast, safety signal learning via conditioned inhibition must include the pairing of the CS+ and CS− in order to test for the active inhibition of threat in the presence of safety. In extinction learning, a new and competing association forms when a cue that was previously associated with threat is repeatedly presented in the absence of threat. In contrast, conditioned inhibition involves associating distinct environmental stimuli (i.e., safety cues) with the nonoccurrence of aversive events (97). Thus, safety cues have the potential to inhibit the expression of fear-related behaviors to cues that signal threat without the competing association involved in extinction (see Table 2 for a comparison of safety signal learning and extinction learning). “Conditioned inhibition” is used hereafter to refer to the process by which a safety cue inhibits fear in the presence of threat, which is tested via summation and retardation test. The relation between conditioned inhibition and general inhibitory control, a core executive function traditionally measured using tasks such as the AX continuous performance task (AX-CPT), stop-signal, go/no-go, and antisaccade tasks in neutral contexts (105), has yet to be investigated. Future research will be important in determining the extent to which these processes are related or distinct.

Safety signal learning has been shown to be effective for reducing fear responding in behavioral studies with rodents (98), nonhuman primates (99), and healthy adult humans (106). In humans, the presence of a safety cue has been shown to reduce physiological correlates of fear, as measured by fear-potentiated startle (106). Most recently, reduced fear-related reactivity in the presence of a safety cue was observed in a cross-species study in adult mice and healthy adult humans using freezing behavior and skin conductance response, respectively (107). Evidence has shown disruption in safety signal learning in adults with PTSD (108; for a review, see 109), suggesting that this process may also be disrupted in anxious individuals relative to nonanxious individuals, although safety signal learning has been explored less in adults with anxiety disorders or following stress. Nevertheless, recent behavioral evidence in rodents suggests that conditioned inhibition of fear via a safety cue may not be as susceptible to the effects of prior stress (e.g., unsignaled foot shocks) as fear extinction (110). These studies suggest that safety cues may effectively reduce fear, even in the presence of a threat stimulus, and may be even more beneficial than typical extinction for individuals with anxiety or a history of stress exposure.

**Neural Mechanisms of Safety Signal Learning via Conditioned Inhibition**

Delineating the neural mechanisms that support safety signal learning is necessary to advance the discovery of neurodevelopmentally informed therapeutics. For example, understanding which neural circuits contribute to safety signal learning could inform how and when in development safety cues might be most effective based on the biological state of the developing brain. Existing knowledge about the neural correlates of conditioned inhibition stems primarily from evidence in rodents and nonhuman primates, with recent evidence in humans contributing to this growing body of literature. Although safety signal learning likely involves integration of information across many regions, the majority of relevant studies have examined single regions in isolation (e.g., via lesions). Based on the literature on fear learning and extinction, the hippocampus, amygdala, and vmPFC have been hypothesized to play a role in safety signal learning via conditioned inhibition.

Given the central role of the hippocampus in contextual fear learning (46, 47, 111) and because the safety cue may provide a context for the CS, the hippocampus has been hypothesized to be involved in conditioned inhibition. Hippocampal projections modulate fronto-amygdala function by supplying information about the degree of threat or safety in the environment (43–45). Through its projections to the vmPFC and BLA (112, 113), the hippocampus influences whether the extinction memory or the original fear memory is behaviorally expressed (44, 46). Safety cues could reduce amygdala reactivity by augmenting prefrontal inputs to the amygdala, specifically those going through the PL cortex or the dACC in humans (45, 107, 111, 114).

Several empirical studies have investigated the role of the hippocampus in safety signal learning (Table 3). Although a pretraining hippocampal lesion has no effect on conditioned...
TABLE 2. Comparison of safety signal learning and extinction learning

<table>
<thead>
<tr>
<th></th>
<th>Safety Signal Learning</th>
<th>Extinction Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>The process through which a stimulus that is overly trained to signal the absence of threat (i.e., the safety cue) reduces, or inhibits, fear in the presence of a threatening cue. Safety signal learning is a class of conditioned inhibition.</td>
<td>The gradual process of fear reduction in classical conditioning that involves repeated exposure to the conditioned stimulus (CS) without the unconditioned stimulus (US).</td>
</tr>
<tr>
<td>Key neural mechanisms</td>
<td>Ventral hippocampal connections with the prelimbic cortex in rodents and the dACC in humans are hypothesized to be involved in inhibiting fear in the presence of safety (107). Dopamine receptors in the basolateral amygdala may also be involved in safety signal learning (208).</td>
<td>The amygdala, vmPFC, and hippocampus are central to fear learning and extinction (23, 31–34). The infralimbic cortex in rodents and the anterior vmPFC in humans inhibit fear expression and store and retrieve the extinction memory (54, 113).</td>
</tr>
<tr>
<td>Age-related differences</td>
<td>Developmental studies are lacking. Hypothesized age-related differences such that safety signal learning may be augmented in adolescence.</td>
<td>Diminished fear extinction learning has been observed during adolescence across species (28, 29).</td>
</tr>
<tr>
<td>Limitations</td>
<td>Safety cues may prevent the generalization of inhibitory learning (146, 153, 159, 160, 162), communicate threat, increase perception of threat, and direct attention away from information that is disinhibitory during exposures (for a review, see 146).</td>
<td>Susceptible to relapse of extinguished fear (extinction memory does not overwrite the original fear association) via the mere passage of time (spontaneous recovery), exposure to a stressor (fear reinstatement), or return to a fear-associated context (fear renewal) (for a review, see 209).</td>
</tr>
</tbody>
</table>

* dACC=dorsal anterior cingulate cortex; vmPFC=ventromedial prefrontal cortex.

inhibition (115, 116), compelling evidence has shown that blocking hippocampal neurogenesis eliminates the behavioral effects of safety signal learning (117) and that ventral hippocampal damage impairs conditioned inhibition (118). Methodologically, these findings suggest that the timing of a lesion, both across development and within a session, may differentially influence behavior. However, the specific mechanism and circuitry through which the hippocampus contributes to safety signal learning remain unknown.

Recent cross-species evidence demonstrated the involvement of the ventral hippocampus in rodents, and the anterior hippocampus in humans, during conditioned inhibition of fear via learned safety (107). Specifically, ventral hippocampal neurons projecting to the PL cortex, targeted using fiber photometry with a retrograde tracer, but not neurons projecting to the IL cortex or BLA, show higher activation during conditioned inhibition, and this activation is associated with lower freezing behavior in mice. A corresponding distinction was observed in humans such that functional connectivity between the anterior hippocampus and dACC—but not hippocampal-anterior vmPFC or hippocampal-amygdala connectivity—was associated with conditioned inhibition. These findings suggest that PL-projecting ventral hippocampal neurons play a role in the inhibition of fear responding in the presence of a safety cue. Importantly, this evidence also suggests that the neural circuitry involved in conditioned inhibition (i.e., ventral hippocampus-PL projections in rodents or hippocampus-dACC connections in humans) differs from that involved in typical extinction (i.e., IL-amygdala projections in rodents or vmPFC-amygdala connections in humans). Previous research had shown that hippocampal inputs to the PL cortex are capable of suppressing, or gating, fear expression (45), further highlighting this pathway as a potential alternative pathway supporting fear reduction. Although further research is needed to test whether PL-projecting ventral hippocampal neurons are causally involved in conditioned inhibition in rodents, and the results in human neuroimaging have yet to be replicated, the parallel results across species strengthen the evidence for the role of this pathway in the active inhibition of fear via learned safety.

The majority of studies have focused on inhibition of the amygdala as a hypothesized mechanism of safety signal learning (Table 3). Emerging findings using in vivo recordings in rodents suggest that a subpopulation of neurons in the BLA, specifically, may be selectively responsive to safety cues, and thus may be involved in the active inhibition of fear through safety cues (119). Notably, amygdala lesion studies have failed to show any effect on summation (120, 121), suggesting that pathways independent of the amygdala can also support conditioned inhibition.

Although the vmPFC plays a central role in discriminating between threat and safety and in extinction learning (e.g., 23, 102), less is known about vmPFC involvement in the active inhibition of fear via safety. Given its dense reciprocal connections with the amygdala and its role in extinction, the vmPFC might down-regulate the amygdala during safety signal learning via conditioned inhibition. Rodent studies suggest that PL and IL subregions of the vmPFC play different roles in this process (122–124). Several studies lesioning the vmPFC (Table 3) have demonstrated mixed findings of disrupted summation (125) or no effect on summation (126, 127). This inconsistency may stem from lesioning the whole vmPFC or its subregions, with some evidence that the IL cortex is selectively involved in conditioned inhibition (125) and other evidence that ventral hippocampal neurons...
<table>
<thead>
<tr>
<th>Authors, Year, Reference</th>
<th>Sample</th>
<th>Region of Interest</th>
<th>Method</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kazama et al. 2012 (120)</td>
<td>Nonhuman primates</td>
<td>Amygdala</td>
<td>Neonatal neurotoxic lesion</td>
<td>Delayed, but not abolished, acquisition of fear; no effect on summation test</td>
</tr>
<tr>
<td>Falls and Davis 1995 (121)</td>
<td>Rodents</td>
<td>CeA</td>
<td>Bilateral electrolytic lesion</td>
<td>No interference with acquisition of conditioned fear inhibition by safety cues</td>
</tr>
<tr>
<td>Sangha et al. 2013 (119)</td>
<td>Rodents</td>
<td>BLA</td>
<td>In vivo single unit recordings</td>
<td>Subpopulation of neurons in the BLA selectively responsive to safety cues</td>
</tr>
<tr>
<td>Ng et al. 2018 (208)</td>
<td>Rodents</td>
<td>BLA</td>
<td>Altering dopamine (D₂) receptors</td>
<td>D₂ receptor alterations (agonist and antagonist) experimentally impair suppression of fear during a safety cue</td>
</tr>
<tr>
<td>Sangha et al. 2014 (125)</td>
<td>Rodents</td>
<td>vmPFC (PL and IL separately)</td>
<td>Inactivation using muscimol/baclofen mixture</td>
<td>Disrupted summation test following IL inactivation</td>
</tr>
<tr>
<td>Rhodes and Killcross 2007 (126)</td>
<td>Rodents</td>
<td>vmPFC (IL)</td>
<td>Lesion</td>
<td>No effect on summation test, but disrupted retardation</td>
</tr>
<tr>
<td>Gewirtz et al. 1997 (127)</td>
<td>Rodents</td>
<td>Large medial PFC region</td>
<td>Lesion</td>
<td>No effect on safety learning, extinction, or retardation</td>
</tr>
<tr>
<td>Kazama et al. 2014 (128)</td>
<td>Nonhuman primates</td>
<td>OFC</td>
<td>Neonatal neurotoxic lesions</td>
<td>No deficits in modulating fear responses in the presence of safety signals</td>
</tr>
<tr>
<td>Sarlito et al. 2018 (138)</td>
<td>Rodents</td>
<td>viOFC</td>
<td>Inactivation using muscimol/baclofen mixture</td>
<td>No effect on discrimination if inactivated during acquisition; inactivation after acquisition and prior to recall impaired discrimination, with the behavioral fear response prevailing</td>
</tr>
<tr>
<td>Fernando et al. 2014 (139)</td>
<td>Rodents</td>
<td>NAc shell</td>
<td>Infusion of dextroamphetamine</td>
<td>Reduced inhibition of avoidance behavior</td>
</tr>
<tr>
<td>Josselyn et al. 2005 (134)</td>
<td>Rodents</td>
<td>NAc</td>
<td>Electrolytic lesion</td>
<td>No impact on acquisition or expression of conditioned inhibition during fear-potentiated startle</td>
</tr>
<tr>
<td>Rogan et al. 2005 (135)</td>
<td>Rodents</td>
<td>Striatum</td>
<td>Electrophysiological recording</td>
<td>Enhanced responses in safety but weakened in fear conditioning, suggesting that safety cues may increase approach behavior</td>
</tr>
<tr>
<td>Christianson et al. 2008 (136)</td>
<td>Rodents</td>
<td>Posterior insula</td>
<td>Excitotoxic lesion</td>
<td>Blocked inhibition via safety cues if lesioned prior to acquisition</td>
</tr>
<tr>
<td>Christianson et al. 2011 (137)</td>
<td>Rodents</td>
<td>Posterior insula</td>
<td>Inactivation using muscimol</td>
<td>Abolished stress-mitigating effects of safety signal learning, but not behavioral control, if lesioned during exposure to uncontrollable stress</td>
</tr>
<tr>
<td>Heldt et al. 2002 (115)</td>
<td>Rodents</td>
<td>Hippocampus</td>
<td>Aspiration lesion</td>
<td>Pretraining lesion did not impair inhibition of conditioned responses, but posttraining lesion did impair performance</td>
</tr>
<tr>
<td>Chan et al. 2003 (116)</td>
<td>Rodents</td>
<td>Hippocampus</td>
<td>Ibotenate lesion</td>
<td>No effect on summation test, though authors note marginal difference such that lesion group responded more to stimulus compound than control group</td>
</tr>
<tr>
<td>Pollak et al. 2008 (117)</td>
<td>Rodents</td>
<td>Hippocampus</td>
<td>Immunohistochemistry and X-irradiation treatment</td>
<td>Increased neurogenesis following safety signal learning; inhibiting neurogenesis prevented the behavioral effects of safety signal learning</td>
</tr>
<tr>
<td>McDonald et al. 2018 (118)</td>
<td>Rodents</td>
<td>Ventral hippocampus</td>
<td>Neurotoxic lesion</td>
<td>Impaired expression of conditioned inhibition such that the lesion group acquired the reversal task more efficiently than the control group</td>
</tr>
<tr>
<td>Meyer et al. 2019 (107)</td>
<td>Rodents and humans</td>
<td>Ventral hippocampus</td>
<td>Fiber photometry and fMRI</td>
<td>Summation test passed in both humans and mice; VH activation related to conditioned inhibition in both humans and mice; activation of VH neurons projecting to PL associated with conditioned inhibition and correlated with freezing in rodents; functional connectivity of VH-dACC associated with conditioned inhibition in humans</td>
</tr>
</tbody>
</table>

a BLA=basolateral amygdala; CeA=central nucleus of the amygdala; dACC=dorsal anterior cingulate cortex; IL=infralimbic cortex; NAC=nucleus accumbens; OFC=orbitofrontal cortex; PFC=prefrontal cortex; PL=prelimbic cortex; VH=ventral hippocampus; viOFC=ventrolateral orbitofrontal cortex; vmPFC=ventromedial prefrontal cortex.
projecting to the PL cortex, and not the IL cortex, are specifically involved in conditioned inhibition (107). While even evidence at the level of subregions has been inconsistent, methods (i.e., muscimol lesion versus fiber photometry) and paradigms of conditioned inhibition differ between these studies and may contribute to mixed results. Furthermore, it should be noted that the study using fiber photometry only measured target-defined ventral hippocampal neuron activity and not activity in the PL or IL cortices themselves. Taken together, these findings suggest an important role of the vmPFC and call for additional research to clarify the precise mechanisms and subregions involved.

Research on conditioned inhibition has focused on several other regions, including the orbitofrontal cortex (OFC) (128), given its hypothesized involvement in inhibitory control (129), behavioral flexibility and reversal learning (130), and encoding stimulus-reward associations (131–133); the striatum, including the nucleus accumbens (NAc) (134, 135), which contributes to motivated behavior and reward processing; and the insula (136, 137), given its widespread connections with the amygdala and involvement in sensory processing and integration (Table 3). Recent studies have provided a more nuanced account of the processes that the OFC may be involved in, including recent lesion-based studies that have suggested that this region is not necessary for inhibition (for a review, see 133). Furthermore, lesion-based research in primates has defined the role of the OFC in value-based decision making, particularly when updating outcome valuations and credit assignment is required (133), as well as in encoding stimulus valuations in conjunction with the amygdala and medial frontal cortex (131; for a review, see 132). In the safety signal learning literature, although a neonatal OFC lesion does not affect summation (128), OFC inactivation after the learning phase disrupts summation (138). Two studies have found that posterior insula lesions block the stress-mitigating effects of a safety cue (136, 137). Inconsistent findings following NAc lesions (Table 3) highlight the need for research to clarify the potential involvement of the NAc in conditioned inhibition (134, 135, 139). These findings suggest that regions outside of the canonical fronto-amygdala fear circuit may also contribute to safety signal learning.

Despite evidence that the hippocampus, vmPFC, and amygdala are involved in conditioned inhibition, and some preliminary evidence for the involvement of the OFC, striatum, and insula, a considerable portion of studies have shown that a lesion to a particular brain region is insufficient to have an impact on safety signal learning (see Table 3). One possible explanation for these inconsistent findings is the differential timing of lesions during development. Many of the lesions were made during the neonatal period, which is followed by extensive brain plasticity. Thus, subjects could have developed compensatory neural pathways that facilitated safety signal learning by the time of testing in adulthood. Furthermore, safety signal learning is likely to involve interactions between brain regions (97). Given widespread connectivity of the amygdala (36, 140), inhibition of amygdala responding could be achieved through inhibitory pathways from local GABAergic interneurons (38), the medial intercalated neurons (141), the IL and PL cortices (38, 142, 143), or input from many cortical regions, including the hippocampus, insula, striatum, and brainstem nuclei (144). Thus, a lesion in a single region may be insufficient to block the effects of safety cues. Future studies in rodents would benefit from circuit-based approaches, such as selective neuronal lesions based on their projections to other regions of interest.

**CLINICAL TRANSLATION: SAFETY CUES IN THERAPY AND THE REAL WORLD**

Although rarely integrated with the neuroscience literature, a rich and growing literature on safety signal learning and safety behaviors (i.e., behaviors that employ a safety cue) exists in clinical science. Patients with anxiety disorders often use safety cues with the goal of reducing fear. Whereas safety cues are often studied in basic science both behaviorally and neurally using standardized paradigms with neutral stimuli (e.g., shapes, sounds, and odors), safety cues outside of the laboratory can take many forms, including people, objects, actions, or mental acts. For example, an individual with panic disorder might carry antianxiety medication, or an individual with social anxiety disorder might enter a new situation only in the presence of their partner. Although conceptualized in basic science research as reducing fear (97), safety cues likely have more complex effects in the everyday lives of individuals with anxiety. Safety cues may reduce anxiety in the short term (94); however, anxious individuals may come to rely on safety cues to function or engage with anxiety-provoking stimuli, such that the cues impede learning that they can tolerate a feared situation. Thus, evidence-based treatments, including CBT, often focus on eliminating patients’ reliance on safety cues (24). In this section, we summarize the literature on clinical research, conducted with adults, on the impacts of safety behaviors in exposure-based therapy, with a particular focus on the potential mechanisms by which safety cues might influence treatment and potential moderators that could guide an understanding of when safety cues might be more helpful rather than detrimental.

**Mechanisms by Which Safety Cues May Enhance Or Interfere With Treatment Outcomes**

In recent years, clinical researchers have proposed that safety cues should be reconsidered in light of theoretical and empirical evidence that their strategic implementation could facilitate fear reduction during exposure-based therapy (145). Empirical evidence that safety cues are detrimental or beneficial is inconsistent (24; for a review, see 146). Recent commentaries and theoretical accounts have highlighted conflicts in conceptualizing safety cues as necessarily detrimental to treatment outcomes (145, 147). A recent meta-analysis determined that the current evidence was too mixed to reach a definitive conclusion about the impact of safety cues.
cues on exposure-based treatment outcomes (148). Delineating the mechanisms by which and the conditions under which safety cues enhance or interfere with symptom reduction will be essential for future research that aims to test the judicious incorporation of safety cues into current treatments. A nuanced examination of the empirical literature on safety cues and safety behaviors in clinical settings may provide insight into whether and how such cues could be carefully integrated into treatment. Of note, existing studies in this realm have focused on adults. Here we first consider mechanisms through which safety cues may enhance or interfere with treatment outcomes, then explore conditions or moderators for efficacy of safety cues in treatment, including their type and timing and developmental stage.

Enhancing treatment acceptability. Exposures are, by definition, an aversive experience for individuals with anxiety, and dropout rates in CBT for anxiety are high, even among adults (149). One way that allowing individuals to engage in safety behaviors may improve treatment efficacy is by making exposures less aversive, and thus enhancing the tolerability of treatment and minimizing patient dropout. Indeed, studies examining treatment acceptability related to safety behaviors have shown that patients rated exposures as more acceptable and reported higher levels of anticipated adherence to treatment (150), as well as lower negative beliefs (i.e., negative misconceptions about a feared object or situation, in this case misconceptions about spiders or about their own reactions during confrontations with a spider) (151) when using safety behaviors. However, not all studies have observed differences in treatment acceptability following use of safety cues in exposures (152, 153).

Facilitating approach behavior. Another proposed benefit of incorporating safety cues into therapy is increasing approach behavior, which is necessary for effective exposure. For example, safety cues could encourage patients to get closer to or accelerate the rate at which they approach a stimulus during exposure (146). Here, too, findings are mixed with regard to the effects of safety cues or behaviors. In some studies, patients have shown increased behavioral approach (150) and accelerated approach behavior (151, 154), as measured using a behavioral approach test (155), when using safety behaviors during exposures. Although there is no evidence to date of detrimental effects on approach behavior, other studies using behavioral approach tests have found that engaging in safety behaviors was not associated with differential approach to the exposure stimuli (152, 156, 157).

Enhancing inhibitory learning. A third potential way that incorporating safety cues may optimize outcomes is by enhancing inhibitory learning. Important theoretical advances have highlighted the role of inhibitory learning in exposure-based therapies for anxiety (24). In fact, an empirical study found that safety cues only reduced startle responses in conditions of perceived threat (as opposed to in the absence of threat), suggesting engagement of top-down regulatory processes associated with inhibition of threat responses (158). Given the implications of inhibitory learning theory for maximizing fear extinction (24), judicious implementation of safety cues may be effective because they involve the active inhibition of fear through safety (i.e., conditioned inhibition). However, it is important to note that safety cues could also interfere with the development of inhibitory associations, and the extent to which safety cues are useful or detrimental may depend on the balance between inhibition and excitation during exposure (104). Further research is needed to elucidate whether and how the inclusion of safety behaviors in exposure-based treatment could effectively augment fear reduction.

Interfering with inhibitory learning. Theoretical models posit that safety cues can interfere with exposure therapy in various ways, including by promoting safety misattributions, disrupting therapeutic information processing, attenuating negative expectancy violation, contextualizing inhibitory learning, and dampening distress tolerance (146). Safety misattributions refer to the hypothesis that when a feared outcome does not occur during exposure, this can be mistakenly attributed to the presence of a safety cue rather than allowing for recognition that the feared outcome itself might be either irrational or tolerable. A few studies have investigated safety misattributions via self-report and have found some evidence for misattribution of safety in exposures including safety behaviors, although, to date, these studies have been limited to nonclinical samples (146, 159, 160). Theories on disruption of therapeutic information processing via the inclusion of safety behaviors include the possibility that the safety behaviors might communicate threat, increase perception of threat, and direct attention away from information that is disconfirmatory during exposures (for a review, see 146). Laboratory studies have found that safety signals, regardless of whether they are inhibitory or excitatory cues, may protect from extinction, and this effect has been attributed to negative expectancy violation (161). Furthermore, the inclusion of safety cues and/or safety behaviors in treatment is tenuous in terms of inhibitory learning theory, which asserts that safety behaviors contextualize inhibitory learning (162). Consistent with this idea, several studies have suggested that safety behaviors prevent the generalization of inhibitory learning. By contrast, other studies using behavioral approach tests have found that engaging in safety behaviors was beneficial to maintaining behavioral approach gains and generalizing extinction (154, 163). Findings on detrimental effects of inclusion of safety cues or behaviors in treatment remain inconsistent and suffer from limitations that include small sample sizes and inclusion of non-treatment-seeking or nonclinical samples (146). Further research is needed to reconcile conflicting findings on the mechanisms by which safety signal learning could affect treatment outcomes.
Conditions Under Which Safety Cues May Enhance Or Interfere With Treatment

Although evidence cautions against the universal implementation of safety cues in treatment, they could potentially augment outcomes if used in specific ways, during specific stages of treatment, or for specific individuals. Identifying the specific conditions under which safety cues facilitate treatment efficacy is an important step in resolving mixed findings in the current literature and informing future efforts to optimize treatments. Here we focus on type of safety cue and developmental stage as two potential moderators for the efficacy of safety signal learning in treatment.

Type of safety cue. In response to inconsistent empirical evidence, a distinction has been proposed between preventive safety cues (i.e., employed prior to an anticipated threatening event to avoid harm or reduce the intensity of the event) and restorative safety cues (i.e., employed following an aversive event with the goal of restoring safety) (164). Restorative safety behaviors have been associated with greater behavioral approach than preventive safety behaviors in an unselected sample of undergraduates who completed a behavioral approach test (165). Moreover, a benign effect of restorative safety behaviors on clinical symptoms and behavioral approach has been shown in the context of exposure-based therapy (164). A recent review (166) classifying studies based on their employment of restorative versus preventive safety behaviors provides additional evidence that restorative safety behaviors may be especially useful. Whereas approximately half of the 23 studies (N=11) using preventive safety behaviors have observed negative outcomes following treatment, all of the studies (N=8) using restorative safety behaviors have found benign or beneficial effects on treatment outcomes. Additionally, empirical studies have found that preventive safety behaviors can block extinction learning. Specifically, individuals who were engaged in a trained safety behavior that could preclude the occurrence of threat (e.g., by making a button press to avoid a shock) continued to show higher reactivity, as measured by pupil dilation (151) and skin conductance level (167), higher self-reported threat expectancy following extinction (167–169), and increased risk for the return of fear (170). One additional study compared the inclusion of preventive versus restorative safety behaviors in treatment for veterans with PTSD and found that inclusion of either type of safety behavior in prolonged exposure therapy was associated with poorer treatment outcomes immediately following therapy, whereas only preventive safety behaviors predicted anxiety at a 3-month follow-up (171). These conflicting findings highlight the need for further research into the precise boundary conditions for preventive versus restorative safety cues in therapy. Although the mechanistic differences between the effects of restorative versus preventive safety behaviors have yet to be examined, restorative safety behaviors might be especially beneficial because they are timed to reduce fear following exposure and thus allow for full confrontation with a core threat and active inhibition of the threat representation (166). As such, they might allow for the in vivo benefits of exposure while also dampening or inhibiting the fear memory trace.

Developmental stage. Most anxiety disorders are diagnosed in adolescence (2), highlighting the need for early intervention. Like in adults, after a course of evidence-based treatment, a substantial proportion of clinically anxious youths continue to meet criteria for an anxiety disorder or experience relapse following initial recovery (11, 12). Some of the factors limiting efficacy may differ between youths and adults, and it has been noted that interventions for youths are largely based on treatment principles studied and implemented in adulthood (27). Moreover, there is limited research guiding clinicians about which types of intervention might maximally benefit anxious patients based on their developmental stage. Thus, delineating when age moderates treatment-related effects is key for optimizing treatment for youths with anxiety (172).

Although less explored, developmental stage (i.e., childhood, adolescence, adulthood) may be an important factor moderating the influence of safety cues on fear reduction. Predictors and moderators of treatment outcome are likely to differ between anxious youths and adults (173–175), and mechanisms supporting the effects of exposure have been less examined in anxious youths than in adults (176). Moreover, the ways in which parental factors relate to child anxiety and the role of parents in treatment present a major difference in treating anxious youths (177–182). However, research on safety signals and behaviors in the context of treatment has been conducted primarily in adults. Here, we propose that based on the neuroscientific evidence on safety signal learning, the judicious use of safety cues in treatment (145) could leverage the inhibitory properties of a conditioned safety cue to be particularly effective for reducing fear in youths with anxiety disorders, and we call for further research on this topic.

**PROPOSED THEORETICAL MODEL OF CLINICAL AND NEUROBIOLOGICAL MECHANISMS BY WHICH SAFETY SIGNAL LEARNING REDUCES FEAR**

Throughout this review, we attempt to draw connections between the literature on neural and clinical mechanisms leading to fear reduction (Figure 1). In the neurobiological literature, evidence includes recent findings on the involvement of the ventral hippocampal–PL pathway—as an alternative to the canonical amygdala–IL pathway involved in fear extinction—as well as the potential involvement of regions including the amygdala, vmPFC, and NAc (97, 107). In the clinical literature, proposed mechanisms for fear reduction via the inclusion of safety cues in treatment include enhancing treatment acceptability, facilitating approach behavior, and altering inhibitory learning (for reviews, see 24, 146, 166). Further research is necessary to disentangle these proposed clinical mechanisms in carefully controlled translational treatment studies. In particular, delineating the conditions

---

**Type of safety cue.** In response to inconsistent empirical evidence, a distinction has been proposed between preventive safety cues (i.e., employed prior to an anticipated threatening event to avoid harm or reduce the intensity of the event) and restorative safety cues (i.e., employed following an aversive event with the goal of restoring safety) (164). Restorative safety behaviors have been associated with greater behavioral approach than preventive safety behaviors in an unselected sample of undergraduates who completed a behavioral approach test (165). Moreover, a benign effect of restorative safety behaviors on clinical symptoms and behavioral approach has been shown in the context of exposure-based therapy (164). A recent review (166) classifying studies based on their employment of restorative versus preventive safety behaviors provides additional evidence that restorative safety behaviors may be especially useful. Whereas approximately half of the 23 studies (N=11) using preventive safety behaviors have observed negative outcomes following treatment, all of the studies (N=8) using restorative safety behaviors have found benign or beneficial effects on treatment outcomes. Additionally, empirical studies have found that preventive safety behaviors can block extinction learning. Specifically, individuals who were engaged in a trained safety behavior that could preclude the occurrence of threat (e.g., by making a button press to avoid a shock) continued to show higher reactivity, as measured by pupil dilation (151) and skin conductance level (167), higher self-reported threat expectancy following extinction (167–169), and increased risk for the return of fear (170). One additional study compared the inclusion of preventive versus restorative safety behaviors in treatment for veterans with PTSD and found that inclusion of either type of safety behavior in prolonged exposure therapy was associated with poorer treatment outcomes immediately following therapy, whereas only preventive safety behaviors predicted anxiety at a 3-month follow-up (171). These conflicting findings highlight the need for further research into the precise boundary conditions for preventive versus restorative safety cues in therapy. Although the mechanistic differences between the effects of restorative versus preventive safety behaviors have yet to be examined, restorative safety behaviors might be especially beneficial because they are timed to reduce fear following exposure and thus allow for full confrontation with a core threat and active inhibition of the threat representation (166). As such, they might allow for the in vivo benefits of exposure while also dampening or inhibiting the fear memory trace.

**Developmental stage.** Most anxiety disorders are diagnosed in adolescence (2), highlighting the need for early intervention. Like in adults, after a course of evidence-based treatment, a substantial proportion of clinically anxious youths continue to meet criteria for an anxiety disorder or experience relapse following initial recovery (11, 12). Some of the factors limiting efficacy may differ between youths and adults, and it has been noted that interventions for youths are largely based on treatment principles studied and implemented in adulthood (27). Moreover, there is limited research guiding clinicians about which types of intervention might maximally benefit anxious patients based on their developmental stage. Thus, delineating when age moderates treatment-related effects is key for optimizing treatment for youths with anxiety (172).

Although less explored, developmental stage (i.e., childhood, adolescence, adulthood) may be an important factor moderating the influence of safety cues on fear reduction. Predictors and moderators of treatment outcome are likely to differ between anxious youths and adults (173–175), and mechanisms supporting the effects of exposure have been less examined in anxious youths than in adults (176). Moreover, the ways in which parental factors relate to child anxiety and the role of parents in treatment present a major difference in treating anxious youths (177–182). However, research on safety signals and behaviors in the context of treatment has been conducted primarily in adults. Here, we propose that based on the neuroscientific evidence on safety signal learning, the judicious use of safety cues in treatment (145) could leverage the inhibitory properties of a conditioned safety cue to be particularly effective for reducing fear in youths with anxiety disorders, and we call for further research on this topic.

**PROPOSED THEORETICAL MODEL OF CLINICAL AND NEUROBIOLOGICAL MECHANISMS BY WHICH SAFETY SIGNAL LEARNING REDUCES FEAR**

Throughout this review, we attempt to draw connections between the literature on neural and clinical mechanisms leading to fear reduction (Figure 1). In the neurobiological literature, evidence includes recent findings on the involvement of the ventral hippocampal–PL pathway—as an alternative to the canonical amygdala–IL pathway involved in fear extinction—as well as the potential involvement of regions including the amygdala, vmPFC, and NAc (97, 107). In the clinical literature, proposed mechanisms for fear reduction via the inclusion of safety cues in treatment include enhancing treatment acceptability, facilitating approach behavior, and altering inhibitory learning (for reviews, see 24, 146, 166). Further research is necessary to disentangle these proposed clinical mechanisms in carefully controlled translational treatment studies. In particular, delineating the conditions
under which safety cues may enhance or interfere with inhibitory learning will be essential to informing interventions.

Here we propose that these neural and clinical mechanisms may interact to enhance or interfere with fear reduction via safety signal learning. Based on the existing clinical literature, safety signal type (i.e., preventative versus restorative safety cues) may moderate the effect of safety signal learning on psychological indices of fear reduction (as in 106, 107), either directly or through the clinical or neural mechanisms. In parallel, based on the existing neurobiological literature, developmental stage, specifically adolescence versus adulthood, may also moderate the effect of safety signal learning on fear reduction. Our conceptual model integrates key clinical and neurobiological factors and highlights the importance of future translational research to test these relations across basic science, including circuitry-to-plasticity tests and differs from fear extinction (see Table 2). Recent cross-species findings highlight the involvement of the ventral hippocampal-PL pathway in safety signal learning (107), and previous research highlights the potential involvement of regions including the amygdala, NAc, and vmPFC (97, 119, 125). Safety cues may affect fear during exposure therapy by enhancing treatment acceptability, facilitating approach behavior, and either augmenting or interfering with inhibitory learning (for reviews, see 24, 146, 166).

These neural and clinical processes may interact to enhance or interfere with fear reduction. Safety signal type (e.g., preventative versus restorative safety cues) (for a review, see 166) and developmental stage (e.g., adolescence versus adulthood) may serve as key moderators of the relation between safety signal learning and fear reduction either directly or through the clinical and/or neural processes. Given marked changes in frontolimbic circuitry from adolescence to adulthood, the neural mechanisms supporting safety signal learning are likely to differ (moderators panel; enlarged in Figure 2) in ways that could lead to differential outcomes in fear reduction, depending on developmental stage. By integrating clinical and neurobiological concepts, this model highlights the importance of future research testing these relations in basic science and treatment studies.

Safety Signal Learning Across Neurodevelopment

Connections between the IL cortex and amygdala in rodents (183–187) and between the vmPFC and amygdala in humans (78–80, 96), which are involved in fear extinction, undergo protracted development during adolescence. Meanwhile, projections between the hippocampus and PL cortex are augmented in adolescence compared with adulthood in rodents (96). The hippocampus has direct projections to the BLA and IL and PL cortices and has been shown to be capable of suppressing fear expression (45, 46, 96, 113). In safety signal learning, the hippocampus may reduce fear responding by augmenting prefrontal inputs to the amygdala (45, 107, 112, 113), with recent cross-species evidence suggesting that this down-regulation may occur specifically through ventral hippocampal (107). Given these potential differences in the circuitry involved in conditioned inhibition via learned safety (i.e., ventral hippocampus-PL or hippocampus-dACC connections) versus in typical extinction (i.e., IL-amygdala or vmPFC-amygdala projections), these mechanisms of fear reduction may operate differently depending on developmental stage.

In addition to existing evidence of increased hippocampal-PL connectivity during the adolescent period in rodents, a separate but related body of literature shows that delivering high-frequency stimulation to the ventral hippocampus can induce long-term changes in the PL cortex, but only after the adolescent period (188–190). Taken together, these findings suggest that there are developmentally dependent changes in the architecture of the hippocampal-PL circuit in focused work with animal models and brain imaging studies with humans, and treatment research, including clinical treatment in tandem with brain imaging to evaluate neural targets identified from basic neuroscience.
adolescence. Future research should investigate the developmental trajectory of the hippocampal-dACC pathway to test whether this circuitry is strengthened in adolescence in humans, which we hypothesize would parallel findings in rodents (96).

Here we propose a theoretical model for the mechanisms through which conditioned inhibition operates based on the developmental state of the implicated neurocircuitry. Given the biological state of the developing brain during adolescence, such that projections between the hippocampus and PL cortex are augmented in rodents (96), activation of hippocampal projections via safety cues may provide an alternate pathway for the optimization of fear reduction (Figure 2). During adolescence, when fear extinction is diminished (29, 191) and frontolimbic circuitry is undergoing substantial changes (79, 192), integrating novel mechanisms of fear reduction that draw on complementary neural circuitry is likely to be especially beneficial. By targeting alternative pathways, as opposed to relying on top-down prefrontal control of the amygdala via traditional extinction, safety signal learning may enhance fear reduction during adolescence, when fear extinction and related vmPFC-amygdala connections are weaker than in adulthood. We hypothesize that safety signal learning via conditioned inhibition would be more efficacious for fear reduction in adolescence, compared with typical extinction. Furthermore, we predict that although adolescents with anxiety disorders would demonstrate poorer safety signal learning relative to nonanxious adolescents, safety signal learning would still be more effective than extinction for anxious youths during this developmental stage. Finally, we predict that greater hippocampal-dACC connectivity during safety signal learning would relate to stronger fear reduction in anxious adolescents. However, no studies to date have examined the neural mechanisms of safety signal learning across development or in anxiety in animal models or humans. Further research is necessary to empirically test this theoretical model (see Box 1), which has important implications for optimizing interventions for anxious adolescents by targeting the state of their developing brains.

When examining the neural mechanisms of safety signal learning, it is important to consider the measures that must be used concurrently to test the efficacy of fear reduction. Particularly in the developmental literature on fear learning, differences in conditioned responding often do not align across different levels of analysis (e.g., self-report, behavioral, physiological) (74, 89, 193). Given issues with reliability, biases, and introspective ability related to self-report measures, particular in adolescents (194), self-report measures should be used in conjunction with behavioral or physiological measures. Behavioral measures, such as the behavioral approach test, and physiological measures, such as pupil dilation and skin conductance response, may be particularly helpful for enabling comparison of findings with the existing literature. Based on previous research showing physiological differences during safety signal learning (106–108), recent evidence that frontolimbic circuitry is involved during safety signal learning in humans (107), and research demonstrating age-related changes in fear learning using physiological measures (74, 89, 193), we expect that differences in safety signal learning related to age and anxiety would be most consistently observed at the physiological and neural levels.

FIGURE 2. Proposed circuitry targeted by safety signal learning during adolescence versus adulthood

Amyg=amygdala; dACC=dorsal anterior cingulate cortex; IL=infraorbital cortex; vmPFC=ventromedial prefrontal cortex. In adolescence, projections from the vmPFC, or the IL in rodents, undergo protracted development (dotted line in adolescence [80, 186]; solid line in adulthood [210, 211]). These projections are involved in fear extinction learning (31, 51), which has been found to be diminished during adolescence across species (29). In parallel, projections between the PL in rodents, or the dACC in humans, and the basolateral amygdala are involved in fear maintenance (39) and have been found to be augmented in rodents during the adolescent period (96). By contrast, projections from the hippocampus to the dACC in humans, or the PL in rodents, are strengthened in adolescence (bolded arrow) relative to adulthood (96). Based on recent findings across species, safety signal learning is thought to involve hippocampal projections to the PL in rodents and the dACC in humans to down-regulate fear responses (107). By targeting an alternative pathway, as opposed to primarily top-down prefrontal control of the amygdala, this approach may optimize fear reduction during adolescence, when fear extinction and related vmPFC-amygdala connections are continuing to develop.

* Amyg=amygdala; dACC=dorsal anterior cingulate cortex; IL=infraorbital cortex; vmPFC=ventromedial prefrontal cortex. In adolescence, projections from the vmPFC, or the IL in rodents, undergo protracted development (dotted line in adolescence [80, 186]; solid line in adulthood [210, 211]). These projections are involved in fear extinction learning (31, 51), which has been found to be diminished during adolescence across species (29). In parallel, projections between the PL in rodents, or the dACC in humans, and the basolateral amygdala are involved in fear maintenance (39) and have been found to be augmented in rodents during the adolescent period (96). By contrast, projections from the hippocampus to the dACC in humans, or the PL in rodents, are strengthened in adolescence (bolded arrow) relative to adulthood (96). Based on recent findings across species, safety signal learning is thought to involve hippocampal projections to the PL in rodents and the dACC in humans to down-regulate fear responses (107). By targeting an alternative pathway, as opposed to primarily top-down prefrontal control of the amygdala, this approach may optimize fear reduction during adolescence, when fear extinction and related vmPFC-amygdala connections are continuing to develop.
Clinical Application of Neurobiological Model

Empirical investigations into treatment for anxious youths should draw not only on the existing literature on conditioned inhibition and clinical use of safety cues, but also on the broader literature on brain development and socioemotional development, to closely inform the timing, nature, and developmental context of the use of safety cues and behaviors. Future developmental studies will benefit from testing systematic incorporation of safety cues in ways that are designed to reduce the aversiveness of exposures, enhance inhibition of fear, and leverage developmentally specific safety cues. Here we provide theoretical examples of how safety cues could be differentially integrated for children and adolescents.

Particular stimuli that could signal safety should be tailored to developmental stage. Early in development, parents are among the most salient stimuli in a child’s life (195). Across species, the mere presence of a parent or a sensory cue related to a parent can reduce fear and stress reactivity in their offspring, referred to as “parental buffering” (196–200). While parents are often involved in an anxious child’s treatment, their role is rarely systematic (201). Integrating parental presence into exposures (followed by gradual fading) could enhance early engagement or acceptability for children. Alternatively, including the parent as a restorative safety cue (165) may be even more effective, potentially allowing children to fully confront the threat cue in the parent’s absence, followed by active inhibition of that threat memory through parental buffering. In this way, we hypothesize that the inhibitory properties of a parent’s presence could be utilized to target an alternate neural pathway for the reduction of fear, which should be tested empirically with treatment-seeking anxious youths using brain imaging during paradigms that manipulate parental presence.

Adolescents are likely to benefit from different forms of safety cues. Naturally salient safety cues in adolescence may include music (202) or the presence of supportive peers (203). Among young adults, pairing threat stimuli with images of social support figures (versus strangers) during extinction has been found to inhibit spontaneous recovery of fear immediately following extinction and during fear reinstatement 24 hours later (204). Integrating a social support figure, such as a close friend, could potentially augment exposure for adolescents. It is important to note that social support figures may be external inhibitors given that they may not have been explicitly trained as signaling safety. However, we propose that using ecologically valid, naturally occurring cues associated with safety (205–207) could provide a valuable starting point to examine the function of safety cues in treatment for anxious youths.

Although judicious use of safety signal learning in treatment for anxious youths has the potential to maximize fear reduction through active inhibition that targets the biological state of the developing brain, much remains unknown about the specific ways in which safety cues could best be incorporated into treatment for youths. Promising avenues for future research include pairing threat and safety cues in time and space during early exposures, using contextual safety cues to enhance fear reduction, restricting safety cues to restorative ones following exposure, or simply removing safety cues more gradually. Given current evidence supporting the use of restorative safety cues in exposures (166), we hypothesize that restricting safety cues to restorative cues will augment the efficacy of exposures in treatment with anxious youths. Systematic tests of specific types and timing of safety cues that are directly informed by knowledge of neural and psychosocial functioning across development will be critical for identifying the most effective ways to augment exposures in anxious youths.

Treatment studies are necessary to investigate whether safety cues implemented in clinical settings rely on the same neural circuitry as those implemented in the literature on conditioned inhibition (i.e., ventral hippocampus-dACC circuitry). These studies should examine the effects of including safety cues in treatment on the brain. For example, clinical trials in youths with anxiety could compare CBT incorporating safety cues and CBT alone and include a brain imaging component before and after treatment. We hypothesize that anxious adolescents would show heightened fear reduction in response to CBT incorporating safety cues, relative to anxious adolescents receiving CBT alone or to anxious adults receiving CBT with safety cues.

As such, these investigations would provide empirical evidence for the proposed theoretical model, and could have important implications for optimizing interventions based on the biological state of the developing brain. Importantly, all of the studies reviewed above have been conducted in adult and young adult patients, highlighting the need for research on safety signal learning and the judicious implementation of safety cues in treatment for anxious youths.

CONCLUSIONS

Anxiety disorders often emerge during childhood and adolescence, yet not all youths benefit sufficiently from current evidence-based treatments. A key feature of anxiety disorders is difficulty regulating fear (16), which may stem from difficulty in learning about or implementing cues that signal safety (99). Behavioral and neuroscientific studies using conditioned inhibition paradigms have shown that safety cues can effectively reduce fear and prevent the onset of new fears in animals (98, 99), and behavioral and neuroimaging studies suggest that safety cues are effective for actively inhibiting fear in humans (106, 107). The hippocampus, vmPFC, amygdala, OFC, NAc, and insula have been proposed as candidate regions involved in conditioned inhibition in animal models. Recent cross-species evidence in humans and rodent models suggests that conditioned inhibition via learned safety may involve connections between the ventral hippocampus and the PL cortex (and connections between the hippocampus and other brain regions).

\[H_2O + CO_2 \rightarrow H_2CO_3\]
fear in youths by targeting the biological state of the developing brain during adolescence (96). Translating these findings on safety cues—e.g., involvement of hippocampal-dorsal anterior cingulate cortex (dACC) connections and further examine neural circuitry as those implemented in the literature on conditioned inhibition (107).

Studies should test whether the use of safety cues is effective for reducing anxiety during childhood and adolescence.

Studies should test different types of safety cues and leverage developmentally specific cues—for example, testing the most advantageous timing (both developmental timing and within-treatment timing) for inclusion of safety cues.

How do the neural circuits involved in safety signal learning change across development?

• Cross-species studies in humans and rodents should test the developmental trajectory of the hippocampal-dACC pathway to replicate previous findings in rodents suggesting that this circuitry is strengthened in adolescence (96) and is involved in safety signal learning.

• Studies that further delineate the developmental timing of hippocampal-fronto-amygdala circuitry changes may be particularly helpful for understanding developmental differences in conditioned inhibition and extinction, as well as informing their potential utility based on the state of the developing brain.

Is safety signal learning differentially effective, relative to extinction learning, at specific stages of development?

• Laboratory studies should compare safety signal learning and extinction learning within subjects across a wide age range from childhood to adulthood.

• Studies should test recall or retention of extinction versus safety signal learning after a delay.

Under which conditions do safety cues facilitate or interfere with symptom reduction?

• Studies that systematically incorporate safety cues into exposures should compare the mechanistic differences between preventive and restorative safety cues.

• Studies should test pairing of threat and safety cues in time and space during early exposures and the gradual removal of safety cues across exposures.

• Studies should include larger treatment-seeking or clinical samples to assess the effects of safety cues on inhibitory learning during exposures and test the potential moderating effect of diagnostic status on fear reduction via safety signal learning.

What psychological or behavioral processes support the effective use of safety cues to inhibit fear?

• Studies should test the effects of safety cues on treatment acceptability, approach behavior, inhibitory learning, or other mechanisms that may mediate fear reduction via safety signal learning.

• Studies should utilize behavioral measures, such as a behavioral approach test, to measure changes in approach behavior during exposures and enable comparison with the existing literature.

To what extent and how do the conditions under which safety cues facilitate symptom reduction change across development?

• Studies should test whether the use of safety cues is judicious use of safety cues may help enhance treatment efficacy under certain conditions, including careful selection of the stages in treatment during which to incorporate and to

and dACC in humans), a pathway that may be strengthened during adolescence (96). Translating these findings on safety cues into the clinical domain may provide a novel approach to reducing fear in youths by targeting the biological state of the developing brain. Here, we propose a neurodevelopmental model of safety signal learning in which the hippocampus plays a central role in down-regulating fear responses through up-regulation of the PL cortex during adolescence.

Existing evidence in the clinical literature suggests that judicious use of safety cues may help enhance treatment efficacy under certain conditions, including careful selection of the stages in treatment during which to incorporate and to
eliminate the use of safety cues. Evidence suggests that the use of restorative safety cues may be particularly effective given that they allow for full confrontation with a core threat and active inhibition of the threat representation. We propose theoretical examples of how safety cues could be differentially integrated for children, such as systematically incorporating parental presence in treatment, and for adolescents, such as the use of music or social support figures. Furthermore, we propose a conceptual model for integrating clinical and neural mechanisms, as well as proposed moderators, for fear reduction via safety signal learning. Although this conceptual model remains largely untested, in this review we formulate hypotheses and recommendations for future translational research in a manner that is consistent with theories and predictions being recorded prior to empirical testing.

Further elucidating the neural mechanisms that support the effective use of safety cues to inhibit fear and how they change across development will be essential to translating research on conditioned inhibition for clinical applications. Integrating empirical findings across these literatures suggests that conditioned inhibition may be especially relevant for children and adolescents with anxiety. Particularly during adolescence, when frontolimbic connections that support traditional fear extinction are still developing (29, 80, 83) and connections between the hippocampus and PL cortex are strengthened (96), the use of safety signal learning may be especially beneficial for optimizing treatment of anxiety based on the biological state of the developing brain. Finally, we call for developmentally informed translational investigations that judiciously incorporate knowledge about the type and timing of safety cues into exposure-based treatments, which have the potential to identify novel ways to maximize fear reduction and optimize treatment outcomes for anxious youths.

**AUTHOR AND ARTICLE INFORMATION**

Department of Psychology, Yale University, New Haven, Conn.

Send correspondence to Dr. Gee (dylan.gee@yale.edu).

Supported by the NIH Director’s Early Independence Award (DP5OD021370) to Dr. Gee, a Brain and Behavior Research Foundation Young Investigator Grant to Dr. Gee, a Jacobs Foundation Early Career Research Fellowship to Dr. Gee, and a National Science Foundation Graduate Research Fellowship Program award (DGE1122492) to Ms. Odirozola.

The authors report no financial relationships with commercial interests.

Received February 28, 2020; revisions received June 17 and July 18, 2020; accepted August 17, 2020.

**REFERENCES**

CONDITIONED INHIBITION FOR FEAR REDUCTION IN THE DEVELOPING BRAIN


32. Pard D, Quirk GJ: When scientific paradigms lead to tunnel vision: lessons from the study of fear. NPJ Sci Learn 2017; 2:6 ()


37. Barad M, Gean P-W, Lutz B: The role of the amygdala in the extinction of conditioned fear. Biol Psychiatry 2006; 60:322–328


42. Laurent V, Westbrook RF: Inactivation of the infralimbic but not the prelimbic cortex impairs consolidation and retrieval of fear extinction. Learn Mem 2009; 16:520–529

43. Fanselow MS: Contextual fear, gestalt memories, and the hippocampus. Behav Brain Res 2000; 110:73–81


46. Fanselow MS, Dong HW: Are the dorsal and ventral hippocampus functionally distinct structures? Neuron 2010; 65:7–19


ity: associations with anxiety, development, and threat. Depress Anxiety 2016; 33:917–926


113. Sierra-Mercado D, Padilla-Coreano N, Quirk GJ: Dissociable roles of prefrontal and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. Neuropsychopharmacology 2011; 36:529–538


119. Sierra-Mercado D, Padilla-Coreano N, Quirk GJ: Dissociable roles of prefrontal and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. Neuropsychopharmacology 2011; 36:529–538


118. McDonald RJ, Balog RJ, Lee JQ, et al: Rats with ventral hippocampal damage are impaired at various forms of learning including conditioned inhibition, spatial navigation, and discriminative fear conditioning to similar contexts. Behav Brain Res 2018; 351:138–151
128. Kazama AM, Davis M, Bachevalier J: Neonatal lesions of orbital frontal areas 11/13 in monkeys alter goal-directed behavior but spare fear conditioning and safety signal learning. Front Neurosci 2014; 8:37