Thesis for doctoral degree (Ph.D.) 2013

Thesis for doctoral degree (Ph.D.) 2013

LEARNING NOT TO FEAR Extinction, erasure, and the recovery of fear memories

Armita Golkar

LEARNING NOT TO FEAR

Extinction, erasure, and the recovery of fear memories

Armita Golkar

Karolinska Institutet

Karolinska Institutet

From THE DEPARTMENT OF CLINICAL NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

LEARNING NOT TO FEAR

Extinction, erasure, and the recovery of fear memories

Armita Golkar

Karolinska Institutet

Stockholm 2013

Published by Karolinska Institutet.

ISBN 978-91-7549-029-8

Printed by
REPROPRINT AB
Stockholm 2013
www.reproprint.se
Gårdsvägen 4, 169 70 Solna

ABSTRACT

Much of the progress in understanding the mechanisms underlying the formation and persistence of fear memories comes from studies of Pavlovian conditioning and extinction. Recently, considerable interest has been turn to strategies that facilitate the development and persistence of extinction. This interest has been particularly fueled by the fact that the findings may have important clinical implications by identifying the conditions during which extinction may permanently prevent the recovery of learned fears. The overall aim of this thesis was to identify the temporal factors that drive fear extinction learning (**Study I**) and to investigate different approaches to preventing the return of fear that occurs after extinction (**Study II-IV**). More specifically, we assessed the effects of initiating extinction training within the consolidation (**Study II**) or reconsolidation (**Study III**) time window and the effects of optimizing safety learning during fear extinction through social observation (**Study IV**).

In **Study I**, we evaluated two critical accounts of extinction by separately manipulating the number of non-reinforced trials and the cumulated non-reinforced exposure time during extinction training. Our data did not support that extinction is driven by the cumulative duration of non-reinforced exposure, but rather the number of trials appeared critical. In fact, many extinction trials with a duration shorter than the acquisition trial duration facilitated extinction learning, but this effect did not predict the recovery of fear.

In **Study II**, we found that extinction training initiated within, but not outside, the consolidation time window yielded less extinction of both fear-potentiated startle and shock expectancy ratings, while selectively preventing the return of fear-potentiated startle during a subsequent reinstatement test. Contrary, in **Study III**, extinction training initiated within the reconsolidation time window did not prevent the recovery of fear, as measured by reinstatement of fear-potentiated startle or skin conductance responses, using either fear-relevant or fear-irrelevant stimuli.

Finally, as an alternative approach to preventing the return of fear, in **Study IV**, we capitalized on the fact that much of what we learn about the environment comes through social forms of learning such as through observation of other individuals.
Therefore, we assessed the effects of vicarious safety learning on the decrement of conditioned fear during extinction training and its effects on the subsequent return of fear. We found that vicarious extinction efficiently reduced conditioned fear responses during extinction and blocked the subsequent return of fear, as measured by skin conductance responses during a subsequent reinstatement test.

In sum, the studies in this thesis demonstrate an intricate relation between extinction learning and the return of fear and highlight that extinction represents a highly complex phenomenon that most probably is determined by multiple factors.

LIST OF PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their roman numerals (**Study I-IV**):

I. **Golkar, A.,** Bellander, M., & Öhman, A. (2012, December 10). Temporal properties of fear extinction - does time matter? *Behavioral Neuroscience.* Advance online publication. doi: 10.1037/a0030892*

II. **Golkar, A.,** & Öhman A. (2012). Fear Extinction in Humans: Effects of Acquisition-Extinction Delay and Masked Stimulus Presentations, *Biological Psychology*, *91*(2).

III. **Golkar, A.,** Bellander, M., Olsson, A., & Öhman, A. (2012). Are fear memories erasable? Reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli. *Frontiers in Behavioral Neuroscience, 6*(80).

IV. **Golkar, A.,** Selbing, I., Flygare, O., Öhman, A., & Olsson, A. (2012). Others as means to a safe end: Vicarious extinction blocks the return of learned fear. *Submitted manuscript*.

ADDITIONAL PUBLICATIONS

Publications by the author from the Department of Clinical Neuroscience which are not included in the thesis:

I. Wiens, S., Peira, N., **Golkar, A**., & Öhman, A. (2008). Recognizing masked threat: Fear betrays but disgust you can trust. *Emotion, 8*, 810-819

II. Lonsdorf, T. B., Weike, A. I., **Golkar, A.**, Schalling, M., Hamm, A. O., & Öhman, A**.** (2010). Amygdala-dependent fear conditioning in humans is modulated by the BDNFval66met polymorphism. *Behavioral Neuroscience*, *124*(1), 9-15213.

III. Peira, N., **Golkar, A.**, Larsson, M., & Wiens, S. (2010). What you fear will appear. Detection of schematic spiders in spider Fear. *Experimental Psychology*, *57*(6), 470-475.

IV. Peira, N., **Golkar, A.**, Öhman, A., Anders, S., & Wiens, S. (2011). Emotional responses in spider fear are closely related to picture awareness. *Cognition & Emotion, 26*(2), 252-260.

V. Lonsdorf, T.B., **Golkar, A.**, Lindström, K.M., Fransson, P., Öhman, A., & Ingvar, M. (2011). 5-HTTLPR and COMTval158met genotype independently gate amygdala activity during passive viewing of angry faces. *Biological Psychology, 87*(1), 106-112.

VI. **Golkar, A**., Lonsdorf, T.B., Olsson, A., Lindstrom, K., Berrebi, J., Fransson, P., Schalling, M., Ingvar, M., & Öhman, A. (2012). Distinct contributions of the dorsolateral and orbitofrontal cortex during emotion regulation. *PLoS ONE, 7*(11).

VII. Lindstrom, KM., Lonsdorf, T.B., **Golkar, A**., Sankin, L., Britton, J., Fransson, P., Öhman, A., & Ingvar, M. 5-HTTLPR genotype influence on right amygdala activation during threat orientation. *Submitted manuscript*.

VIII. Lindström, B., Mattson-Berglund, I., **Golkar, A**., & Olsson, A. In your face: Risk of punishment enhances cognitive control and error-related activity in the corrugator supercilii muscle. S*ubmitted manuscript*.

CONTENTS

LIST OF ABBREVIATIONS

B	Basal nucleus
BLA	Basolateral amygdala
BOLD	Blood-oxygen-level-dependent
CE	Central nucleus
CR	Conditioned response
CS	Conditioned stimulus
CS+	Conditioned stimulus coupled to US
CS-	Conditioned stimulus never coupled to US
DCS	D-cycloserine
fMRI	Functional magnetic resonance imaging
FPS	Fear-potentiated startle
GABA	Gamma-aminobutyric acid
IL	Infralimbic cortex
ITC	Intercalated cells
ITI	Inter-trial interval
KDEF	Karolinska directed emotional faces
LA	Lateral nucleus
mPFC	Medial prefrontal cortex
NMDA	N-methyl-D-aspartate
PFC	Prefrontal cortex
RET	Rate-expectancy theory
SCR	Skin conductance response
US	Unconditioned stimulus
vmPFC	Ventromedial prefrontal cortex

1 INTRODUCTION

1.1 FEAR LEARNING AND EXTINCTION

Learning to predict danger is fundamental to survival. Pavlovian conditioning is an exemplar of this type of learning, and enables the organism to form associations between threatening events and preceding innocuous cues (e.g., sounds, smells). The functional significance of this mechanism is that it allows the organism to anticipate danger and prepare appropriate defense systems to cope with an impending threat in advance of its actual occurrence (Öhman & Mineka, 2001). Although such evolved defense systems serve adaptive purposes, persistent conditioned responding to events that no longer predict danger can develop into pathological anxiety. In fact, conditioned fear is regarded as one of the primary mechanisms in the etiology of fear-related anxiety disorders (Mineka, & Zinbarg, 2006) and Pavlovian fear conditioning represents the leading model to study the neural and behavioral mechanisms through which such fears are acquired and stored.

In clinical practice, fear-related anxiety disorders are effectively treated by cognitive behavioral therapy (Barlow, 2002), which derives its effectiveness from the repeated exposure to the feared object in the absence of aversive outcomes. The experimental analogue of exposure therapy is represented by the process of fear extinction, during which the expression of a previously learned fear response is weakened through repeated exposures to the fear-eliciting cue when it no longer predicts aversive consequences. The inability to extinguish fear responses when they are no longer appropriate is a hallmark of many anxiety disorders. Consequently, the objective of most behavioral therapies is to reduce resistance to extinction learning and promote the formation of new associations that eliminate the fear response. Although adopting the principle of fear extinction has proved effective in treatment, still a considerable number of patients are not helped and others suffer from relapse episodes during which extinguished fears return (Foa, 2000; Rachman, 1989). Therefore, one way to understand how exposure treatment can be optimized to reduce the risk of relapse is by understanding the basic processes that govern extinction learning and the mechanisms through which previously extinguished fears reappear. As such, extinction represents an important model both for developing knowledge of basic learning processes and for bridging experimental findings to applied settings.

The overall aim of this thesis was to identify the temporal factors that drive fear extinction learning (**Study I**) and to investigate different approaches to preventing the return of fear that occurs after extinction (**Study II-IV**). The first part of this thesis (Introduction) will start with a brief overview of the neural properties of fear learning and extinction to highlight the existence of a well conserved, evolutionarily shaped neural network centered on a small structure in the medial temporal lobe of the brain, the amygdala. Then I will review some of the basic behavioral properties that characterize fear extinction and introduce the associative learning framework from

which most theoretical accounts of extinction derive. Finally, I will describe different strategies that have been employed to study how the return of learned fear can be prevented. Specifically, I will focus on strategies in which extinction learning interferes with the consolidation or reconsolidation of fear memories and strategies that focus on optimizing safety learning during fear extinction.

1.2 NEURAL PROPERTIES

Fear is as an unpleasant, often strong emotional state elicited by anticipation or awareness of danger and is associated with rapid instinctive responses related to avoiding and preparing for conflict (Öhman, 2000). The expression of fear is characterized by a common psychophysiological response pattern including potentiation of the startle reflex, increases in skin conductance response (SCR), blood pressure and heart rate acceleration (Globisch, Hamm, Esteves, & Öhman, 1999). The observed psychophysiological response pattern suggests that these fear-related processes are mediated by the center of the brain's fear network; the amygdala. The amygdala is a small structure composed of a collection of anatomically and functionally distinct nuclei located within the temporal lobe (Pitkanen, Savander, & LeDoux, 1997). A pivotal role for the amygdala in mediating the acquisition and expression of fear has been well established mainly based on studies using Pavlovian fear conditioning protocols (Davis, 2003; LeDoux, 2000). Pavlovian fear conditioning represents a basic form of learning to predict danger. It reflects the process by which an initially neutral stimulus (conditioned stimulus; CS) acquires behavioral relevance when paired with an innately aversive stimulus (unconditioned stimulus; US) in a manner that allows the organism to learn that the CS predicts the occurrence of the US. As a result of learning this CS-US relation, the CS acquires the ability to elicit defensive responses that are normally elicited in the presence of danger. These defensive responses include behaviors such as freezing, autonomic and endocrine responses such as heart rate acceleration and hormonal release, as well as the expression of reflexes such as the fear-potentiated startle (FPS).

The underlying neuroanatomical circuitry has been well described in rodents with the use of lesions to or pharmacological inactivation of specific nuclei within the amygdala. Two subregions within the amygdala are particularly important for fear conditioning: the *basolateral complex* (BLA), which includes the lateral (LA) and basal (B) nuclei, and the *central nucleus* (CE). Briefly, information about the CS and US seems to converge in the LA that sends its output to the CE. The CE in turn controls the expression of conditioned responses (CR) through descending projections to other regions, including projections to the hypothalamus that are important for mediating autonomic responses, and projections to structures in the brainstem that regulate the behavioral expressions of fear (Davis, 1992; Fendt & Fanselow, 1999; LeDoux, 2000; Maren, 2001). In humans, studies of fear conditioning have replicated many of the basic findings derived from studies in rodents (see LeDoux, 2000; Phelps & LeDoux, 2005 for reviews). Thus, both lesion studies (e.g. Bechara et al., 1995; LaBar, LeDoux, Spencer, & Phelps, 1995) and functional imaging studies in humans have been supportive of a key role of the amygdala in the acquisition of conditioned fear (e.g. Buchel, Morris, Dolan, & Friston, 1998; LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998; see Sehlmeyer et al., 2009 for a review), suggesting that the underlying fear circuit has been well conserved across species.

Given the central role of the amygdala in mediating fear learning, the amygdala has also been implicated in the extinction of fear (Davis, Walker, & Myers, 2003), which is commonly studied by repeatedly presenting the CS in the absence of its associated US. In rodents, the study of amygdala involvement in fear extinction has not been as straightforward as the study of its role in fear conditioning. This is partly due to the fact that classical approaches such as electric or neurotoxic lesions of the amygdala are not useful for the study of fear extinction since the amygdala is required for both the acquisition and the expression of fear itself (see LeDoux, 2000; Maren, 2001, for reviews). The question has however been addressed with the use of alternative approaches showing that amygdala activity changes during extinction in both rodents (Quirk, Repa, & LeDoux, 1995; Repa, Muller, Apergis, Desrochers, Zhou, & LeDoux, 2001; Rogan, Staubli, & LeDoux, 1997) and humans (Gottfried & Dolan, 2004; Knight, Smith, Cheng, Stein, & Helmstetter, 2004; LaBar et al., 1998; Milad, Wright, Orr, Pitman, Quirk, & Rauch, 2007; Phelps, Delgado, Nearing, LeDoux, et al., 2004) and that extinction is associated with specific molecular processes within the amygdala (for reviews see Herry et al., 2010; Myers & Davis, 2007).

In rodents, the majority of studies have specifically targeted the BLA as a candidate site mediating extinction learning. The rationale for this has mainly been based on fear conditioning studies indicating that the BLA, and specifically the LA, show properties of learning-related neural plasticity (for reviews see Blair, Schafe, Bauer, Rodrigues, & LeDoux, 2001; Maren, 1999). During conditioning, LA neurons increase their firing rate in response to the CS and this increase in CS-elicited activity has been shown to be reversed during extinction training (Quirk et al., 1995; Repa et al., 2001). This reversal, manifested as a decrease is spiking rate, is displayed by most LA neurons and is correlated with a reduction of the CR. Interestingly, not all LA neurons display this reversal pattern, but some maintain high spike firing throughout extinction training (Repa et al., 2001). This sustained firing pattern has also been suggested to be context-dependent, as LA neurons fire specifically in response to the CS when the CS is presented in a context outside the extinction context (Hobin, Goosens, & Maren, 2003).

On a molecular level, the neural plasticity underlying extinction learning seems, at least in part, to be mediated by glutaminergic N-methyl-D-aspartate (NMDA) receptors within the amygdala (Walker & Davis, 2002). Thus, consistent with the role of NMDA receptors in mediating neural plasticity in different forms of learning and memory (Martin, Grimwood, & Morris, 2000), including fear conditioning (Maren & Fanselow, 1995; Miserendino, Sananes, Melia, & Davis, 1990), local infusion of a NMDA receptor antagonist into the rat BLA prior to extinction has been shown to dose-dependently block extinction of conditioned fear (e.g. Falls, Miserendino, & Davis, 1992). Others have demonstrated that blockade of NMDA receptors after extinction training results in high CR during a subsequent extinction recall test (Santini, Muller, & Quirk, 2001; Suzuki, Josselyn, Frankland, Masushige, Silva, Kida, et al., 2004), suggesting that NMDA-receptors are involved in the consolidation of extinction memory. Moreover, the opposite strategy (i.e. improving the activity of the same receptor) has been shown to facilitate extinction. Thus, the partial NMDA-receptor

agonist D-cycloserine (DCS) administered either systemically or directly into the rat BLA before extinction training dose-dependently enhanced extinction of the FPS reflex (Walker, Ressler, Lu, & Davis, 2002) and of conditioned freezing (Ledgerwood, Richardson, & Cranney, 2003). Moreover, consistent with the proposed role of NMDA receptors in the consolidation of extinction memory, DCS exerts facilitating effects when given up to 3 hr after extinction training (Ledgerwood, Richardson, & Cranney, 2005).

Although the amygdala is evidently involved in mediating extinction, amygdala processes alone do not seem to be sufficient to explain all neural aspects of extinction. Rather, extinction processes seem to depend on interactions between the amygdala and cortical areas such as the medial prefrontal cortex (mPFC) (Maren & Quirk, 2004; Sotres-Bayon, Bush, & LeDoux, 2004). The amygdala and the mPFC are reciprocally connected in both rodents (McDonald, Mascagni, & Guo, 1996; Vertes, 2004) and primates (Ghashghaei & Barbas, 2002) allowing for functional interactions between these structures. An early demonstration of the involvement of the ventromedial PFC (vmPFC) in extinction was provided by Morgan et al (1993) who showed that rats with vmPFC lesions induced prior to conditioning displayed impaired fear extinction but retained the ability to acquire conditioned fear. As an extension of the well documented effects of response perseveration after damage to the PFC (see Sotres-Bayon et al., 2004 for a review), Morgan and colleagues (1993) suggested that the observed impairments in extinction represented a form of emotional perseveration; an inability or failure to inhibit fear to a CS that has ceased to signal threat. Subsequent work has argued that lesions to the vmPFC in rats do not cause a general impairment in extinction learning but rather cause specific deficits in the ability to remember extinction. Thus, rodents with lesions to the infralimbic cortex (IL) of the vmPFC are unable to recall extinction when tested 24 hr after extinction training (Quirk, Russo, Barron, & Lebron, 2000). Alternative approaches to lesion studies have provided converging evidence for the role of vmPFC in extinction recall. Thus, inactivating agents infused directly into the IL in rodents impair extinction retrieval (Burgos-Robles, Vidal-Gonzalez, Santini, & Quirk, 2007; Santini et al., 2004; Sierra-Mercado, Corcoran, Lebron, Milad, & Quirk, 2006), and conversely, direct stimulation of the IL enhances extinction retrieval (Milad & Quirk, 2002; Milad, Vidal-Gonzalez, & Quirk, 2004). Currently, there are two alternative models of top down regulation of the amygdala by the vmPFC in rodents. Briefly, because projections from the vmPFC to the amygdala are largely excitatory (Smith, Pare, & Pare, 2000), the inhibition exerted by the vmPFC is thought to involve activation of inhibitory interneurons located within the amygdala. Thus, the first model suggests that excitatory projections from the IL of the vmPFC inhibit the BLA projections to the CE via activation of local inhibitory GABAergic interneurons located within the BLA (Grace & Rosenkranz, 2002). The second model, however, posits that the IL excites inhibitory intercalated (ITC) projection neurons situated between the BLA and CE and these projection neurons in turn inhibit the CE output (Pare, Quirk, & LeDoux, 2004).

Human functional magnetic resonance imaging (fMRI) studies using fear conditioning protocols report that mPFC activity changes during different phases of extinction (Gottfried & Dolan, 2004; Kalisch, Wiech, Hermann, & Dolan, 2006; Milad, et al., 2007; Phelps et al., 2004) and in line with the work from non-human animals, both structural (Milad, Orr, Pitman, & Rauch, 2005) and functional data (Kalisch et al., 2006; Milad et al., 2007; Phelps et al., 2004) indicate that the vmPFC is particularly involved in the recall of extinction. The first of these fMRI studies on extinction recall in humans (Phelps et al., 2004) reported increased activity in the vmPFC during an extinction recall session that occurred 24 hr after extinction training. Specifically, this activity was localized to the subgenual anterior cingulate cortex, which is a subregion of the vmPFC that has been proposed to be the human homologue of the IL in rodents (Kim, Somerville, Johnstone, Alexander, & Whalen, 2003). Consistent with animal models of top-down control (Quirk, Likhtik, Pelletier, & Pare, 2003; Rosenkranz, Moore, & Grace, 2003), the functional association between activity in the amygdala and vmPFC during recall has been suggested to reflect PFC activation of local inhibitory interneurons within the amygdala that suppress the expression of fear (Milad et al., 2007).

Moreover, consistent with neural models of extinction in rodents (Moustafa et al., 2013), the functional network supporting the recall of extinction in humans seems to include the hippocampus, which a structure located adjacent to the amygdala in the medial temporal lobe. In the context of fear learning and extinction, the hippocampus is involved in assembling contextual and temporal information about the environment in which learning occurs (for reviews see Bouton, 2004; Bouton, Westbrook, Corcoran, & Maren, 2006). Thus, when introducing a contextual shift between fear conditioning and extinction the observed mPFC-amygdala activity during the recall of extinction has been associated to an increased activity in the hippocampus (Kalisch et al., 2006; Milad et al., 2007), suggesting a functional connectivity between the mPFC, the hippocampus, and the amygdala (Milad et al., 2007). Moreover, the hippocampus has also been suggested to play a fundamental role in the inhibition of anxiety-related responses in post-traumatic stress disorder (Rauch, Shin, & Phelps, 2006).

Taken together, human neuroimaging studies have been consistent with non-human animal models of extinction learning, suggesting that the neural processes underlying fear extinction and recall have been conserved across species. The wealth of data from non-human animals coupled with the evidence of a shared network offers a unique opportunity to derive specific and well-informed hypotheses about the neural and behavioral properties of extinction learning in humans.

1.3 BEHAVIORAL PROPERTIES

Most theories of extinction learning originate from an associative framework. Collectively, they assume that conditioning involves the formation of representations of the CS and the US and the contexts in which they occur, as well as about the relationships between these stimuli and the contexts. These learned relationships between representations are described as associations. Although specific theories differ in their assumptions regarding the factors that govern the formation of associations, most theories explain the acquisition of CR as resulting from the formation of excitatory associations between representations of the CS and US. The presentation of the CS both activates the CS representation and, indirectly, the US representation via its association with the CS representation, which consequently triggers the CR. Thus, CR reflects the strengthening of the connections between the internal representations of the CS and the US, which are commonly referred to as the associative strength of the CS.

An important distinction concerns learning about predictive relations from learning about contiguous relations, i.e. the temporal pairing of events, a distinction that dates back to an elegant paper by Rescorla (1967) in which he proposed that simply pairing the CS and the US does not sufficiently explain learning in Pavlovian conditioning. Procedurally, fear conditioning involves a specific temporal relation between a CS and an aversive US such that the CS precedes the occurrence of the US. It has long been known that breaking this contiguous relation by increasing the temporal interval between the offset of the CS and the onset of the US severely retards learning (Yeo, 1974). This observation has fueled the idea that the temporal relation between the CS and US is critical for learning. However, learning about predictive relations involves learning about the causal relationship between events (Dickinson, 1980; Rescorla, 1988). This view implies that the CS and the US must be correlated so that the CS provides unique information about the occurrence of the US. Thus, according to this associative framework, the mechanisms that govern conditioning depend on both contiguity, the CS and US must occur together in time, and contingency, they must occur in a predictive relationship.

Within this associate framework, “unlearning” accounts describe extinction as resulting from the destruction of the excitatory association between the CS and the US so that the CS representation fails to activate the US representation and consequently does not trigger the CR. Although the idea that extinction can cause unlearning has been pervasive and was originally incorporated in influential theories assuming that new learning destroy old learning (Rescorla & Wagner, 1972), most current accounts represent extinction as a form of new learning. In fact, the notion of extinction as a form of learning has prevailed for several decades and the idea was present already in the seminal work of Pavlov (1927) on the basis of his experiments with conditioned salivation in dogs. More recently, one of the most influential “new learning” accounts have been put forward by Bouton (1993), who has proposed that extinction reflects learning of a new CS-no US association that competes with the

original, excitatory CS-US association. This "new-learning" account is supported mainly by four post-extinction phenomena during which previously extinguished CRs recover. These phenomena suggest that under specific experimental conditions, the CR can return. Collectively, they highlight a critical role of temporal and contextual factors in determining extinction (see Bouton, 2002 for a review).

The most well studied recovery effect is *spontaneous recovery,* which was first documented by Pavlov (1927) who noted that the extingushed response to a CS can spontaneously recover with the passage of time, suggesting that the decrease in CR during extinction is a transient effect (see Rescorla, 2004 for a review). A number of explanations for this effect have been put forward, including those that focus on attentional processes (Robbins, 1990), and those focusing on a failure to retrieve the original CS-US association due to a swith of temporal context (Brooks & Bouton, 1993). Taken together, the spontaneous recovery effect suggest that the extinction memory is less stable than the acquisition memory beacuse it is more affected by lapse of time.

A second source of recovery is represented by the effect of *renewal,* during which extingushed CRs reappear when tested outside of the context in which extinction training ocurred, suggesting that extinction memory is more senstive to contextual changes than is the acquisition memory. Thus, if subjects acquire a CR in context A and are extingusihed in a different context B, then responding to the CS will only be reduced in the extinction context B but not if re-exposure occurs in the acquistion context A, or in a novel context C (Bouton & Bolles, 1979a). As argued by Bouton (see Bouton, 2004 for a review), context seems to play a modulatory role, because what is learned is not that the CS is *not* predictive of the US but raher that the CS is not predictive of the US in a particular context.

Reinstatement represents a third source of recovery and involves the reapperance of extingushed CRs after unsignaled presentations of the US and was first described by Pavlov (1927) and later confirmed by Rescorla (Rescorla & Heth, 1975) who made two important observations. First, reinstatement was cue specific, because the response did not generalize to a neutral CS. Second, the reinstated response was not due to a local sensitization effect, since it was evident 24 hr after the unsignaled US presentations. Subsequent research in both rodents and humans has demonstrated that reinstatement only occurs if the unsignaled US is presented in the same context as the reinstatement test (e.g. Bouton & Bolles, 1979b; Bouton & King, 1983; LaBar & Phelps, 2005).

Finally, the forth phenomenon supporting that extinction does not erase the original learning comes from *reacquistion* experiments, which have shown that introducing additional CS-US pairing after extinction results in relearning of the original CS-US association at a faster rate than during initial learning. This suggests that the original fear memory was partly "saved" throughout extinction training. One explanation for this effect implies that the first CS-US presentation during reacquisition resets the

acquisition context and thereby reactivates the CS–US memory (see Bouton, 2002 for a review).

Whereas research in non-human animals has put considerable effort into understanding the mechanisms underlying the return of conditioned fears, mechanistic approaches to explaining these effects in humans remain scarce. Nevertheless, there are numerous studies demonstrating the presence of these effects in humans (see Hermans, Craske, Mineka, & Lovibond, 2006 for a review) and that have demonstrated some of their fundamental properties. These include that reinstatement only occurs in a group of subjects re-exposed to the US compared to control group not re-exposed to the US (Hermans et al., 2005; Norrholm et al., 2006) and that renewal of CR occurs if extinction is conducted in a different context than acquisition and testing (i.e. in a so-called ABA design) but does not occur in the absence of a context switch (i.e. in a AAA design) (Vansteenwegen et al., 2005).

It is important to note that although most accounts of extinction learning are associative in nature, non-associative mechanisms such as habituation-like procesess have been suggested to at least partly influence extinction (Kamprath & Wotjak, 2004; Robbins, 1990). Habituation refers to the decrease in responsivness to a stimulus as a result of repeated presentations or after a prolonged time of exposure (Thompson & Spencer, 1966). The idea that habituation mechanisms participate in extinction is not new, and was already incorporated in some early theories (Pearce & Hall, 1980; Rescorla & Heth, 1975). More recently, McSweeney and Swindell (2002) argued for the role of habituation in extinction by highlighting that extinction and habituation share several fundamental properties. These include that both show spontaneous recovery and stimulus specificity, although there are several additional properties that are distinct to extinction learning, such as the demonstration of faster reacquistion. Additional evidence that extinction and habituation share common mechanisms comes from molecular research implicating the endogenous cannabinoid system in both processes (Kamprath et al., 2006; Marsicano et al., 2002). Thus, athough there is abundant evidence supporting that extinction involves the formation of a new associative memory, it seems likely that extinction is influenced by multiple factors.

1.4 WHAT CAUSES EXTINCTION?

Perhaps the most fundamental question regarding extinction concerns what actually drives the waning of conditioned fear responding during extinction training and that defines the process in terms of its effects on learned fear. However, surprisingly little is known about the processes that govern the decrease in conditioned fear responding. Like much of the general literature on fear learning and extinction, the available work mainly comes from studies in non-human animals. These studies have provided ample evidence that time is integral to the acquisition and expression of conditioned fear, but it is still unclear exactly which temporal characteristics are critical in determining the decrease in CR during extinction. Procedurally, extinction involves both a progressive increase in the number of non-reinforced CS trials and a progressive increase in the duration of non-reinforced exposure to the CS, raising the question of which of these temporal properties that critically determine extinction.

In the context of conditioned fear responses, an early study by Shipley (1974) in rats set out to determine whether fear extinction was governed by the number of extinction trials or the duration of exposure to the CS. This was accomplished by manipulating the duration of the CS so that either a short (25 s) or a long (100 s) CS predicted the onset of the shock. Shipley (1974) reported that the duration of the extinction trial did not predict extinction as long as animals received an equal amount of CS exposure. Based on these findings, Shipley proposed that the extinction of conditioned fear is a function of the total amount of non-reinforced exposure to the CS. However, the interpretation of this study suffers from several methodological constraints, such as the introduction of a contextual shift between conditioning and extinction and extinction re-test, and the fact that responses were only assessed in an extinction re-test session that occurred at differed temporal intervals after the final extinction trial.

According to the most influential associative learning model, originating in the work by Rescorla and Wagner (1972), extinction is assumed to reflect the weakening of the influence of the CS–US association such that repeated non-reinforced CS trials results in a reduction in the associative strength of the CS. More formally, the model states that the associative strength (V) that accumulates to a CS on a particular trial is a function of the discrepancy, or predictive error, between the actual outcome of the conditioning trial (λ) and the expected outcome of the conditioning ($\sum V$). The expected outcome of the conditioning trial is the summed associative strengths of all CSs present on that trial. Excitatory conditioning occurs when the actual outcome of the trial exceeds the expected outcome (i.e. $\lambda > \sum V$) whereas no conditioning occurs when the actual and expected outcomes are the same (i.e. $\lambda = \sum V$). In these terms, extinction occurs when the expected outcome ($\sum V$) exceeds the actual outcome (λ) so that the discrepancy ($\lambda - \sum V$) is negative. Thus, learning is represented as a change in associative strength and these changes in associative strength occur in response to events, i.e. on a trial-to-trial basis. As such, the Rescorla-Wagner model predicts that extinction will progress as a function of the number of non-reinforced trials, but at least in its original formulation,

did not predict how changing temporal parameters, such as the inter-stimulus-interval between the CS and US or CS trial duration, would affect the progress of extinction (cf Brandon, Vogel, & Wagner, 2002).

More recent developments of the Rescorla-Wagner model (1972), including a class of real-time models, have more explicitly accounted for the temporal phenomena that are associated with fear acquisition. In contrast to the traditional trial-based model, real-time models depart from the assumption that learning occurs continuously across a trial rather than on a trial-to-trial basis (e.g. Schmajuk & Moore, 1989; Sutton & Barto, 1981). Consequently, according to these models, there can be multiple prediction errors generated throughout one single trial. Among these models, the so called componential trace models (Brandon et al., 2002) assume that the CS is a compound cue that is composed of multiple successive cues. These include temporal and sensory cues, which can independently acquire associations with the US. Such models predict that extinction learning requires non-reinforced presentations of the original acquisition CS duration. Thus, training subjects with a CS of a given duration, but extinguishing them with a shorter CS duration, will result in little long-term extinction, because non-reinforced exposure to the training CS duration never occurred. In contrast, lengthening the CS duration from acquisition to extinction is predicted to have negligible effects on extinction, since non-reinforced presentations to the learned CS duration will still occur. Still other computational models (e.g. Grossberg & Schmajuk, 1991) predict that extinction will be unaffected by changing the CS duration because the expectations of US delivery is timed from the onset of CS. According to this view, subjects encode information about when in time the US will be delivered in relation to the onset of the CS, but they do not necessarily encode information about the duration of the CS. Thus, changing the duration of the CS is not predicted to affect extinction.

A radically different approach is taken by so called time-based models, which argue that associative learning theory fails in providing an adequate account of the temporal properties of conditioning and extinction. Rather, such models describe the acquisition of CR in terms of learning of temporal intervals and the duration and rate of events. In these terms, extinction begins when the animal decides that the US rate in the presence of the CS has changed. Perhaps the most influential time-based model has been formalized in rate-expectancy theory (RET) proposed by Gallistel & Gibbon (2000). According to RET, extinction is determined by the cumulated duration of non-reinforced CS presentations. More specifically, the model predicts that extinction reflects a decision process based on the ratio between the cumulated CS duration after the last US and the expected US waiting time, and as such, the number of extinction trials is irrelevant. Rather, as the cumulated amount of non-reinforced exposure increases, the ratio of these variables approaches a criterion at which CR stops.

With more recent data, it still remains inconclusive whether trial-based models such as the Rescorla-Wagner model (Rescorla & Wagner, 1972) or time-based models such as RET (Gallistel & Gibbon, 2000), best explain the decrease in CR during extinction. Collectively however, studies from non-human animals have suggested that both

extinction (Haselgrove & Pearce, 2003) and extinction re-test performance (Drew, Yang, Ohyama, & Balsam, 2004; Plendl & Wotjak, 2010) are sensitive to changes in CS duration. The general approach in these studies (Drew et al., 2004; Haselgrove & Pearce, 2003) has been to condition animals with a fixed CS –US interval and then extinguish them with a CS duration that was either longer, shorter, or the same as the CS duration used during acquisition. Changing the CS duration between acquisition and extinction was shown to facilitate the decrease in CR, but when re-exposing animals to the acquisition CS duration after extinction, the animals extinguished to a CS duration different from the acquisition duration displayed the most recovery of CR (Drew et al., 2004). These data suggest that the effectiveness of the extinction training depended on the degree of dissimilarity between the acquisition and extinction CS duration.

Against this background, the aim of **Study I** was to disentangle the contribution of cumulated number of trials and exposure time to extinction and the recovery of fear. Moreover, we investigated whether changing CS duration from acquisition to extinction testing could facilitate extinction learning and the recovery of fear.

1.5 CAN EXTINCTION ERASE FEARS?

Although "new learning" and "unlearning" theories of extinction often are presented as mutually exclusive, it has also been acknowledged that both mechanisms may contribute to extinction (e.g. Delamater, 2004). One of the critical observations supporting this view is that the recovery of CR after extinction is often partial relative to the level expressed by a control group that did not receive extinction training. This partial recovery of the CR suggests that, at least to some degree, erasure does occur. If extinction under certain conditions can cause erasure of learned fears, it opens an avenue to investigate how the expression of once learned fear memories can be prevented. The challenge has been to identify the experimental conditions during which fear memories can be erased and prevented from returning.

In 2006, Myers and colleagues (Myers, Ressler, & Davis, 2006) revived interest in the idea of erasure mechanisms by suggesting that different mechanisms mediate extinction depending on the temporal delay between fear acquisition and extinction. In a series of studies in rodents, they reported that extinction that started shortly (10 minutes) after fear acquisition did not result in reinstatement, renewal or spontaneous recovery of the FPS reflex, but that these hallmarks of extinction were present when extinction training started 24 -72 hr after acquisition. Thus, the authors suggested that erasure mechanisms might preferentially be invoked when extinction training is initiated shortly after fear acquisition, whereas inhibitory learning accounts for the mediation of extinction once the fear memory has been stabilized (Myers & Davis, 2007). But what is the mechanism whereby the timing of extinction can modulate the expression of fear memory?

From a theoretical point of view, the differences between immediate and delayed extinction can be understood in the context of consolidation theory. Consolidation refers to the process whereby memories progressively become more stable and is thought to serve an adaptive function by allowing endogenous systems, such as the adrenergic system, to strengthen memories of emotionally arousing events (McGaugh, 2004). The term memory consolidation was first proposed more than 100 years ago in the seminal work of Müller & Pilzecker on the acquisition and retrieval of verbal information in humans, in which they demonstrated that the memory of newly learned information was disrupted by learning that occurred shortly after the original learning (reviewed in Lechner, Squire, & Byrne, 1999). They proposed that the processes governing new memories initially exist in a labile state where they are sensitive to disruption, but progressively become stable and resistant to the same disruptive factors. During the last century, studies in a wide range of species and learning tasks have shown that consolidation of new memories can be disrupted by several types of interference. These include interference with molecular/cellular process such as inhibition of protein synthesis and the expression of certain genes, as well at the system-level such as interference induced by brain trauma (for a review see McGaugh, 2000).

One of the key questions involves the timing of interference because the precise time course of consolidation remains unclear. This is partly due to the fact that the different phenomena that have been labeled consolidation occur at varying time scales. *Systems consolidation* refers to processes that are involved in the reorganization of brain regions involved in the retrieval of a memory, so-called explicit or declarative memories, and operate on the scale of month to years, and in humans, even decades. *Cellular consolidation processes,* on the other hand, operate on the scale of hours and refer to the initial set of cellular/molecular processes, such as activation of specific genes and protein synthesis that are recruited to support the local strengthening of the synapses. Experimental work in non-human animals, such as rodents, has shown that such synaptic consolidation of conditioned fear memory can be disrupted by intra-amygdala infusions of a protein-synthesis inhibitor called anisomycin (Kwapis, Jarome, Schiff, & Helmstetter, 2011; Schafe & LeDoux, 2000; Wilensky, Schafe, Kristensen, & LeDoux, 2006). These studies suggest that the consolidation time window during which disruption of conditioned fear memories is possible opens a few minutes after training and lasts up to 6 hr after training.

Similar temporal constraints of long term memory formation have also been described in the context of reconsolidation, which is the process whereby previously consolidated memories can be reactivated and again rendered sensitive to disruption (Nader, Schafe, & LeDoux, 2000b; Sara, 2000). According to one dominate view, this recurrent window of vulnerability serves an adaptive function by representing a mechanism whereby old memories can be updated with new information (Alberini, 2005). The phenomenon of reconsolidation has been documented since the late 1960s (Misanin, Miller, & Lewis, 1968), but it took approximately 30 years until the broad interest in reconsolidation mechanisms revitalized with the demonstration that consolidated fear memories can be reactivated and again rendered sensitive to disruption (Nader, Schafe, & LeDoux, 2000aa). Since then, although not demonstrated to be ubiquitous (but see Lee, 2009 for an alternative perspective), memory reconsolidation has been documented in several different species, from invertebrates to rodents and humans, and in different types of learning tasks including those that target hippocampus-dependent spatial memory, aversive memories, and human episodic memory (Alberini, 2005; Dudai & Eisenberg, 2004; Nader & Hardt, 2009). The vast majority of the recent reconsolidation studies have been conducted in non-human animals using Pavlovian fear conditioning paradigms. In general, the procedure includes establishing a fear memory by exposing the animal to predictive CS-US pairings in a classical fear conditioning paradigm. A day later, after allowing the memory to be fully consolidated into long-term storage, the fear memory is reactivated by a single presentation of the CS that is presumed to initiate the reconsolidation process.

As the definition implies, reconsolidation and consolidation share several features, such that both processes are sensitive to interference by protein synthesis inhibitors, beta-adrenergic receptor antagonists, and new learning; but they also show properties that are distinct to one process or the other, such as the dependence on partly different brain areas (Alberini, 2005). Interestingly though, the critical time window during which

these processes are sensitive to disruption are partly overlapping. Thus, in rodents, the critical reconsolidation time window has been suggested to open minutes after reactivation and to last at least 1 hr (Monfils, Cowansage, Klann, & LeDoux, 2009) to eventually be completed after 6 hr (Duvarci & Nader, 2004; Nader et al., 2000a). Thus**,** intra-amygdala infusions of a protein synthesis inhibitor (anisomycin) immediately, but not 6 hr, after reactivation of the fear memory significantly reduced conditioned fear responses at a later retention test (Nader et al., 2000a). Subsequently, Debiec & LeDoux (2004) showed that both systemic and intra-amygdala injection of the beta-adrenergic receptor antagonist propranolol blocked reconsolidation, whereas enhancing noradrenergic activity in the amygdala have been demonstrated to enhance reconsolidation and strengthen fear memory (Debiec, Bush, & LeDoux, 2011). Interestingly, the effects of manipulating noradrenergic activity with systemic propranolol administration have been extended by Kindt, Soeter, and Vervliet (2009) in a human fear conditioning paradigm. They reported that the recovery of conditioned FPS in humans could be blocked with pre-reactivation administration of propranolol while sparing the declarative memory of the CS-US relationship (i.e., shock expectancy). Similar beneficial effect of propranolol on reconsolidation have been reported in patients with post-traumatic stress disorder (Brunet et al., 2008), suggesting that pharmacological disruption of fear memory reconsolidation may be an effective intervention for reducing fear and anxiety.

A related line of research has demonstrated that replacing pharmacological treatment (i.e. propranolol) with extinction training yields similar results. Thus, extinction training initiated within, but not outside, the critical reconsolidation time window has been shown to attenuate or block the return of conditioned fear, as first described in rodents (Monfils et al., 2009) and later extended to humans (Schiller et al., 2010) using skin conductance responses (SCR). In the study by Schiller et al (2010), subjects were first fear conditioned to two different CSs. A day later, the experimental group received one non-reinforced CS reminder trial followed by extinction training within the reconsolidation time window, whereas two control groups received either the reminder trial and extinction training outside of the reconsolidation time window or no reminder trial at all but extinction training only. The authors showed that 24 hr later, the expression of fear, as measured by a renewal test, was selectively abolished in the experimental group, suggesting that extinction training initiated within, but not outside, of the critical time window erased the expression of fear memory. Importantly, the study by Schiller et al (2010) showed that the effect of extinction training initiated within the reconsolidation window persisted 1 year later, as measured by the absence of expressed fear during a reinstatement test.

If extinction training initiated either within the consolidation or the reconsolidation time window can cause a permanent erasure of the fear memory, the clinical implications for the treatment of anxiety disorders could be profound. However, there are several issues that require further attention.

First, the interpretation of the study by Myers and colleagues (Myers et al., 2006), suggesting that extinction training conducted immediately after fear acquisition erased the return of fear, is complicated by the fact that previous studies in humans have demonstrated a significant return of fear following an immediate extinction procedure (e.g. Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Hermans et al., 2005; LaBar & Phelps, 2005; Schiller et al., 2008). Nevertheless, the findings by Myers et al (2006) do raise the question of whether there are quantitative differences between immediate and delayed extinction that support the view that fear memories are more easily disrupted immediately after they have been acquired than after they have undergone consolidation. Second, in the context of reconsolidation mechanisms, if interference is to prove effective to treat clinical fears, the original findings by Schiller et al (2010) require replication and extension to increase their clinical applicability. Against this background, the overall aim of **Study II** and **III** was to further study the effects of extinction training initiated within the consolidation (**Study II**) or reconsolidation time window (**Study III**) on the return of conditioned fear.

No doubt interfering with consolidation or reconsolidation processes represents promising avenues to erase the expression of learned fears, but there are several issues complicating the applicability of these strategies. In a clinical context, preventing the return of fear by interfering with consolidation of fear memories is complicated by the fact that the "original" fears are often learned days or years before treatment is initiated, i.e. there is seldom a chance to interfere with consolidation since the memories have already been consolidated. In this perspective, interfering with reconsolidation holds greater promise because it capitalizes on the dynamic properties of memory formation and maintenance. On the other hand, interfering with reconsolidation is constrained by several boundary conditions. Thus, previous work has shown that interference with reconsolidation is temporally graded, such that recent memories are more sensitive to disruption than more remote memories (Frankland et al., 2006; Suzuki, Josselyn, Frankland, Masushige, Silva, & Kida, 2004) and that the temporal dynamics of reconsolidation are dependent on the strength of the acquired memory (Suzuki, Josselyn, Frankland, Masushige, Silva, & Kida, 2004; Wang, Alvares, & Nader, 2009). Given these constrains, it is valuable to explore alternative strategies to preventing the return of fear. One alternative strategy to disrupting the formation or reformation of the acquired fear memory is to strengthen the formation of the safety memory, i.e. enhancing extinction learning per se.

Recent progress in determining the molecular processes underlying extinction have given rise to a number of pharmacological agents that have been shown to facilitate extinction in non-human animals (See Kaplan & Moore, 2011 for a recent review). Thus, pharmacological agents targeting the glutaminergic (Ledgerwood et al., 2003; Mao, Hsiao, & Gean, 2006; Parnas, Weber, & Richardson, 2005; Walker et al., 2002; Woods & Bouton, 2006; Zushida, Sakurai, Wada, & Sekiguchi, 2007), the monoaminergic (Cain, Blouin, & Barad, 2004; Morris & Bouton, 2007; Ponnusamy, Nissim, & Barad, 2005), as well as the endocannabinoid and glucocorticoid systems (i.e. Chhatwal, Davis, Maguschak, & Ressler, 2005; Yang, Chao, & Lu, 2006) (see

Mariano de Bitencourt, Pamplona, & Takahashi, 2013 for a review) have shown to exert extinction-facilitating effects in preclinical trials in rodents. Of these, the partial NMDA receptor agonist DCS has provided the most promising results and has been shown to augment therapeutic outcomes in humans. Thus, DCS given in conjunction with exposure therapy has been reported to result in significant clinical improvement in patients with acrophobia (Ressler et al., 2004) social phobia (Hofmann, Pollack, & Otto et al., 2006), obsessive-compulsive disorder (Kushner et al., 2007) and specific phobia (Guastella, Dadds, Lovibond, Mitchell, & Richardson, 2007; but see also Guastella, Lovibond, Dadds, Mitchell, & Richardson, 2007). Although additional research is needed to determine the clinical value of drugs such as DCS, the progress within this field holds promise for the development of new treatment strategies that may augment the efficacy of current exposure-based behavioral therapies for anxiety disorders. However, in spite of this excitement, due to the side effects that accompany most drugs and to the development of drug tolerance that render treatment ineffective with time, alternative behavioral approaches are preferred over pharmacological manipulations.

In clinical settings, one behavioral approach to enhance safety learning during exposure is offered by observational or *vicarious* safety learning, which has long been exploited as a part of exposure treatment of phobias. In such treatment, the therapist – acting as a learning model –approach and interact with the phobic stimulus before the phobic individual is directly exposed to it (Seligman & Wuyek, 2005). The principle underlying such participating treatment can be explained by one of the most influential theories in psychology; social learning theory. Much of the development of this theory is ascribed to the work of Albert Bandura, who through a series of studies in the 1960s investigated how children learned through observing the behavior of others (Bandura, 1977). In one of these so-called "Bobo doll" experiments (Bandura, Ross, & Ross, 1961), one group of 3-6 year old children were exposed to an adult learning model that acted aggressively towards a large inflatable plastic doll (hence the name Bobo doll). During a subsequent test, children that had observed an aggressively acting model were more likely to imitate the aggressive behavior as compared to a group of children exposed to a passive model or a group that were not exposed to a model at all, and adding incentives increased the children's tendency to express aggressive behavior. Subsequent work by Bandura and his colleagues (Bandura, Grusec, & Menlove, 1967) demonstrated that children's phobic responses to dogs could be extinguished by observing another child's fearless interaction with dogs after a series of modeling sessions. Although promising, this and other early behavioral studies (Bandura et al., 1967; Hill, Liebert, & Mott, 1968; Ritter, 1968) suffered from several methodological limitations, such as unsatisfactory control conditions. Also, these early as well as the few existing more recent studies (Gilroy, Kirkby, Daniels, Menzies, & Montgomery, 2001) only included phobic participants, thereby limiting the generalization of the results, and exclusively relied on behavioral measures.

Although more recent data on vicarious learning of safety are scarce, considerably more experimental work has been reported on vicarious learning of fears (Bandura & Rosenthal, 1966; Berger, 1962; Hygge & Öhman, 1978; Mineka & Cook, 1993;

Mineka, Davidson, Cook, & Keir, 1984; Olsson & Phelps, 2004). In a typical setup, the participant observes another person - the learning model, undergo a conditioning procedure during which the model starts displaying emotional reactions coupled to the presentation of a formerly neutral stimulus. Vicarious fear learning is inferred from the participant's emotional responses to the presentation of this stimulus, because the participant has no direct aversive experience coupled to the presentation of that stimulus. An early demonstration of vicariously learned fear in humans was offered by Hygge & Öhman (1978). In that study, participants were exposed to another person, a confederate acting as the learning model, expressing fear reactions in response to two different classes of stimuli; a fear relevant stimulus such as a snake, and a fear-irrelevant stimulus, such as a flower. Observing a model expressing fear paired with the presentation of these stimuli was sufficient to elicit physiological fear responses (i.e. SCR) in the observer. Moreover, this vicarious learning of fear was more pronounced towards fear-relevant than fear-irrelevant stimuli. Similar findings were subsequently reported by Mineka and her colleagues in a series of studies on vicarious fear conditioning in monkeys (Mineka & Cook, 1993; Mineka, Davidson, Cook, & Keir, 1984) in which they showed that laboratory-reared monkeys could acquire strong and persistent fears to snakes after observing another monkey behaving fearfully with snakes.

More recent studies have shown that acquisition of fear through social observation shares several features with directly acquired fear (Askew & Field, 2008; Hooker, Verosky, Miyakawac, Knight, & D'Esposito, 2008; Hygge & Öhman, 1978; Kelly & Forsyth, 2009; Olsson, Nearing, & Phelps, 2007; Olsson & Phelps, 2004). For instance, Olsson & Phelps (2004) showed that participants expressed equivalent levels of conditioned fear responses during an observationally learned fear conditioning task as during a direct or instructed fear conditioning task. Moreover, both directly and observationally acquired fears were expressed when stimuli were presented under masked conditions that precluded participants conscious awareness of the learned fear stimuli. Subsequent studies have shown that the expression of fear acquired directly or observationally both involve the amygdala (Olsson, Nearing & Phelps, 2007), highlighting a partly shared neural network (see Olsson & Phelps, 2007 for a review).

Still, little is known about the processes that govern learning to attenuate conditioned fears through social observation. This is striking given that much of the information about our environment is learned from other individuals, which is an ability that has been well conserved across species (Olsson & Phelps, 2007). Against this background, the aim of **Study IV** was to develop an experimental design that allowed us to investigate the contribution of vicarious extinction learning to attenuating previously learned fears and to investigate whether this type of safety learning could prevent the return of such fears.

2 AIMS

The overall aim of this thesis was to investigate the processes that govern fear extinction learning and to investigate different approaches to preventing the return of fear that occurs after extinction. To achieve this overarching aim, we specified the following objectives:

- To evaluate the temporal characteristics governing fear extinction learning (**Study I**).
- To assess the effects of interfering with consolidation of fear memory with extinction training initiated within the consolidation time window on the recovery of conditioned fear (**Study II**).
- To assess the effects of interfering with reconsolidation of fear memory with extinction training initiated within the reconsolidation time window on the recovery of conditioned fear (**Study III**).
- To asses the effects of enhancing safety learning through social observation during extinction training on the recovery of conditioned fear (**Study IV**).

3 METHODS

3.1 PARTICIPANTS

All participants (N= 203) were screened for life-time psychiatric disorders and current or past psychopharmacological medication. Only healthy, non-medicated participants were included in the final samples of **Study I-IV**. Before participation, all participants gave written informed consent and were paid for their participation at the conclusion of the experiments.

3.2 STIMULI

We used male faces expressing fearful (**Study I-III**), or angry (**Study IV**) facial expressions from the Karolinska Directed Emotional Faces (KDEF; Lundqvist, Flykt, & Öhman, 1998) as CSs. In the context of fear learning, fearful and angry facial expressions belong to a class of stimuli often referred to as fear-relevant. Previous research has shown that conditioned fear to fear-relevant stimuli, such as images of angry or fearful expressions or images of spiders or snakes, share several features with phobic fears. These features include resistance to extinction, fast acquisition rate and insensitivity to verbal information (see Öhman & Mineka, 2001 for a review). Coupled with the fact that the feared object in clinical fears are more often fear-relevant than fear-irrelevant (Öhman and Mineka, 2001), our rationale was that using fear-relevant stimuli as CSs would increase the clinical relevance of our findings. As a comparison, we used fear-irrelevant (colored squares) CSs in experiment 2 of **Study III**.

3.3 VISUAL MASKING

In **Study II**, we used a technique known as visual masking. Procedurally, visual masking, or backward masking more specifically, involves a brief presentation of a target picture that is followed by a masking picture. Given the proper temporal parameters and technical requirements this procedure results in participants reporting that they only see the masking picture but not the preceding target (Enns & Di Lollo, 2000; Wiens & Öhman, 2007). Previous research has shown that conditioned fear to fear-relevant stimuli can survive masking (Morris, Öhman, & Dolan, 1998; Öhman & Soares, 1993), implying that, under some circumstances (i.e. when stimuli are fear-relevant), explicit awareness of the CS-US contingencies is not necessary for the expression of conditioned fear (Esteves, Dimberg, & Öhman, 1994).

3.4 PARTIAL REINFORCEMENT SCHEDULES

We used partial reinforcement schedules in all fear acquisition protocols presented in this thesis (**Study I-IV**), i.e. the proportion of the CS+ trials that was followed by a shock varied from 50% to 82%. Although the partial reinforcement schedule

complicates the computation performed by the participant, our rationale was primary based on previous human conditioning studies (e.g. Dunsmoor, Bandettini, & Knight, 2008; Kindt & Soeter, 2011; LaBar et al., 1998; Phelps, Delgado, Nearing, & LeDoux, 2004; Schiller et al., 2010) in which partial reinforcement schedules are commonly employed to study the recovery of fear. This has primary been done to slow extinction learning, because extinction learning generally occurs rapidly in humans (LaBar et al., 1998), and to slow re-extinction learning during the subsequent recovery test.

3.5 PSYCHOPHYSIOLOGICAL MEASUREMENTS

3.5.1 Fear-potentiated startle (Study I-III)

The startle reaction is a fast defensive reflex that is elicited in response to a sudden and intense stimulus (startle probe) such a loud noise (acoustic startle reflex) or a light flash. The startle reflex has been well conserved across species and represents one of the most reliable indices of fear mobilization (Lang, Davis, & Öhman, 2000). In humans, the first and most reliable index of the startle reflex is the startle *blink* reflex. It reflects the contraction of the orbicularis oculi muscle in response to a sudden sensory stimulus and the magnitude of this contraction can be quantified by electromyographic recordings using surface electrodes attached underneath the eye. The blink reflex is the component of the startle response most commonly used as an index of CR in human fear conditioning preparations, and within this context, it is commonly referred to as the fear-potentiated startle (FPS). The basis of the FPS is that the startle blink reflex is potentiated when the individual is in an aversive or fearful state and the magnitude of the startle reflex has been shown to be directly related to negative stimulus valence (Hamm & Weike, 2005). On a neural level, it reflects the influence of direct and indirect connections from the amygdala to the primary startle reflex pathway in the brain stem (Davis & Whalen, 2001) and appears to index a basic, affective level of fear conditioning (Öhman & Mineka, 2001).

In the context of fear conditioning paradigms in humans, which commonly use visual stimuli as CSs, the startle probes are often acoustic, such as a sudden burst of white noise known to elicit the startle blink reflex. These probes are presented in the presence of the CS that is predictive of the US (the CS+), a control CS (the CS-), and during the intervals between the CSs (inter-trial interval, ITI). The index of the CR is inferred from the difference in the magnitude of the startle reflex elicited in the presence of the CS+ compared to the CS- and the ITIs. Importantly, the FPS is an invaluable tool in translational research on fear because the human startle response is comparable to the whole-body startle response that is expressed in other non-human animals such as rodents, it generates a non-zero baseline against which the CS-elicited responses can be related, and is mediated by a well-characterized neural system (Davis, 1997). Despite these advantages, the use of FPS to index conditioned fear has drawbacks. The main drawback is related to the aversive nature of the acoustic startle probe which might interfere with or retard fear learning or extinction (Myers et al., 2006). In the context

of human fear conditioning studies, technical constrains have limited the concurrent measurement of FPS and Blood-oxygen-level-dependent (BOLD) responses in the amygdala, although such studies are starting to emerge (van Well, Visser, Scholte, & Kindt, 2012).

3.5.2 Skin conductance response (Study III-IV)

The skin conductance response (SCR), also known as the electrodermal response, represents one of the oldest measurements in the history of psychophysiological research and dates back to the 1880s. SCRs reflect the phasic increase in skin conductance that occurs in response to physiologically arousing stimuli, such as negative or positive pictures, and are modulated both by stimulus novelty and intensity as well as by attentional processes (Öhman, 1979). SCRs to emotional or significant stimuli in humans are predominately under sympathetic control and reflect the activity of eccrine sweat glands that are abundant in the palm of the hands and sole of the feet. Sympathetic activation of the sweat glands lowers the resistance of the skin which results in an increase in conductance. In humans, SCRs are routinely measured by a pair of electrodes attached to the distal phalanges of the index and middle finger of the hand. Notably, the measurement of SCR is inherently event-related due to the lack of a non-zero baseline.

In the context of human fear conditioning, SCR represents the most widely used index of CR and is commonly inferred from increased SCRs in the presence of a CS that is predictive of the US (the CS+) relative to SCRs in the presence of the control stimulus (CS-). The difference between these measures is referred to as differential SCR. One of the strongest advantages with assessing SCR lies in the fact that it offers a non-intrusive physiological index (i.e. does not require the presentation of an eliciting sound or light) of physiological activation. Moreover, SCR represents the most common concurrent index of CR in human fMRI studies, and as such has provided correlational data relating changes in SCRs to changes in amygdala activity during fear conditioning (e.g. Cheng, Knight, Smith, & Helmstetter, 2006; Knight, Nguyen, & Bandettini, 2005; LaBar et al., 1998).

4 OVERVIEW OF THE STUDIES

4.1 STUDY I

Background and objectives

Procedurally, extinction involves both a progressive increase in the number of non-reinforced CS trials and a progressive increase in the duration of non-reinforced exposure to the CS, but it is unclear which of these temporal properties that critically determines extinction. Conceptually, two distinct models have been invoked to account for the decrease in CR that accompanies extinction. Rescorla and Wagner (1972) postulated that extinction is critically determined by the number of non-reinforced exposure trials, whereas Gallistel & Gibbon (2000) proposed that the decisive event is the cumulated exposure time to the non-reinforced CS elapsed after the last CS reinforcement. Against this background, the objective of **Study I** was to evaluate the temporal characteristics of extinction in a human differential fear conditioning study by independently manipulating the number of trials and the cumulated exposure time.

Methods

Sixty-two students at Karolinska Institutet (19 men) participated in the study in one single session. Prior to starting the experiment, participants were randomly assigned to one of four different experimental groups. As outlined in Table 1, the groups differed in terms of the number of non-reinforced trials, the amount of cumulated non-reinforced exposure, and the number of non-reinforced presentations of the acquisition duration.

The experiment consisted of three different phases; fear acquisition, extinction and reinstatement testing. First, all participants were subjected to a fear conditioning protocol in which they watched a presentation of two different pictures depicting fearful male faces from the Karolinska Directed Emotional Faces (Lundqvist et al., 1998), which served as CSs. Each CS was presented for 6 s and a white fixation cross was shown on a black background during the inter-trial intervals (ITIs; 12-15 s). The presentation of one CS (the to-be CS+) was coupled to an electric shock (the US), which was delivered to the lower ventral right arm on six of the nine CS+ presentations, whereas the other stimulus (the CS-) served as a control and was never coupled to a shock (nine non-reinforced presentations). Which face that served as CS+ or CS- was counterbalanced between participants. During extinction, participants received one of four different extinction protocols that varied in terms of the amount of cumulated non-reinforced exposure, the number of non-reinforced trials, and the number of non-reinforced presentations of the acquisition duration (see Table 1). At the end of the extinction session, all participants underwent a standard reinstatement procedure during which they received three unsignaled presentations of the US followed by four non-reinforced presentations of each CS with the original 6-s acquisition duration.

To assess conditioned fear responses, we measured the magnitude of the FPS. To elicit the startle reflex, startle probes (50 ms bursts of 95-dB[A] white noise with a near instantaneous rise time) were presented to the participants through headphones 4-5 s after stimulus onset on one-third of all CS presentations throughout the experiment and on an equal number of ITIs.

The basic rationale for the analysis was as follows. If the total duration of non-reinforced CS trials is critical, as suggested by time-based models, the groups should not differ after receiving the same duration of cumulative non-reinforced CS exposure. If however, the total number of non-reinforced CS trials is critical, as suggested by trial-based models, the groups should not differ after receiving the same number of CS trials.

Table 1. Extinction parameters

Group (CS duration x No. of trials)	No. of trials	Cumulated non-reinforced exposure (s)	No. of non-reinforced presentation of acquisition CS duration
6 s x 12	12	72	12
6 s x 24	24	144	24
3 s x 24	24	72	0
12 s x 12	12	144	12

Results & Conclusions

Our data did not support that extinction is driven by the cumulative duration of CS exposure. Thus, equating groups based on the cumulated exposure time (72 s) but allowing the number of received extinction trials to vary, did not result in an equivalent decrease in CR. Rather, the group that received the fewest number of trials, and the longest CS trial duration, showed least extinction. Equating groups based on the total number of exposure trials (12 trials) on the other hand, and allowing the cumulated exposure time to vary between groups did not reveal any significant group differences in CR, supporting that the number, but not duration, of non-reinforced trials seems to determine the decrease in CR during extinction. However, these effects were not retained in the data when extinction progressed beyond the first 72 s of CS exposure. In fact, with an extended extinction protocol, extinction progressed most readily with many exposure trials (24 trials) with a CS duration shorter (3 s) than the acquisition CS duration, although this effect did not predict the subsequent recovery of fear. This finding is in line with other work (Craske et al., 2008; Prenoveau, Craske, Liao, & Ornitz, 2012), suggesting that the decrease in CR during extinction training is not predictive of the subsequent expression of CR.

4.2 STUDY II

Background and objectives

According to contemporary accounts (Bouton, 1993), extinction represents an inhibitory learning process during which non-reinforced presentations of a CS result in the formation of a new memory that competes with the original excitatory memory acquired during conditioning. This view is mainly supported by the fact that under certain circumstances, the expression of extinguished fears can recover. This account has however been challenged by a series of experiments in rodents (Myers et al., 2006) suggesting that extinction might be mediated by different mechanisms depending on the temporal delay between fear acquisition and extinction. Thus, extinction training initiated shortly after fear acquisition was reported to abolish the return of fear, whereas extinction training initiated with a sufficient temporal delay (24 -72 hr) left the recovery of fear intact. The main aim of **Study II** was to evaluate these effects in a human differential fear conditioning paradigm.

Methods

Thirty-three students at the Karolinska Institutet participated in the study. Prior to starting the experiment, participants were randomly assigned to two different groups; one group received acquisition and extinction training during one single session and one group received extinction training 24 hr after acquisition. All participants were first fear conditioned to four different pictures depicting fearful male faces from the Karolinska Directed Emotional Faces (Lundqvist et al., 1998), which served as CSs. Each CS was presented for 6 s and a white fixation cross was shown on a black background during the ITIs (12-15 s). Two of the CSs (the to-be CS+s) were coupled to an electric shock (the US), which was delivered to the lower ventral right arm on nine of each of the two CS+ presentations, whereas the other two stimuli (the CS-s) served as controls and were never coupled to a shock. To facilitate discrimination between the faces, the faces were presented on two different background colors. One pair of CSs (one CS+ and one CS-) was always presented on a blue background and the other CS pair was always presented on a yellow background. Which face that served as CS+ or CS- within each pair was counterbalanced between participants as was the coupling between CS pair and background color.

During extinction, all participants were presented with 12 non-reinforced presentations of each of the four CSs. To explore whether varying the acquisition-to-extinction interval would have an effect on reinstatement in the absence of CS-US contingency learning during extinction, we including masked CS trials that precluded CS-US contingency learning during extinction. Thus, for each participant, one CS pair (one previously reinforced CS+ and one CS-) was immediately masked by a neutral face (33 ms CS immediately followed by a 6-s mask presentation) while the other CS pair was

presented non-masked (33 ms neutral mask immediately followed by a 6-s CS presentation). The critical effect achieved by backward masking, in contrast to merely very short stimulus presentations, is that is allows controlled exposure to a short stimulus but precludes its access to conscious processing, thus preventing association of the CS with US omission (Enns & Di Lollo, 2000). At the end of the extinction session, all participants underwent a standard reinstatement procedure during which they received three unsignaled presentations of the US followed by four non-reinforced and non-masked 6-s presentations of each of the four CSs. Finally, to verify that the parameters used during the extinction session efficiently masked the CSs, all participants completed a post-experimental forced choice recognition task. During this task, participants were exposed to the same face-mask pairings as during the experiment and their task was to indicate which face the first picture depicted.

To assess conditioned fear responses, we measured the magnitude of the FPS. To elicit the startle reflex, startle probes (50 ms bursts of 95-dB[A] white noise with a near instantaneous rise time) were presented to the participants through headphones 4-5 s after stimulus onset on two-thirds of the CS presentations during acquisition and extinction testing, and during three-fourths of the reinstatement testing, and during an equal number of ITIs per phase. Additionally, we collected trial-by-trial shock expectancy ratings throughout all experimental phases by instructing participants to move a lever to the right side if they expected to receive a shock or to the left side if they did not expect to receive a shock.

Results and Conclusions

Our results provide partial support for the hypothesis that the recovery of conditioned fear responses varies with the temporal delay between acquisition and extinction training. Thus, for the non-masked CS pair, we observed the expected recovery of fear following delayed extinction training, as assessed by a CS+ specific increase in CR during reinstatement testing, which was not evident for the immediately extinguished CS+. In contrast, reinstatement of shock expectancy ratings occurred independently of the acquisition-to-extinction delay as both extinction groups showed a significant increase in shock expectancy ratings to the non-masked CS+ from end extinction to reinstatement testing. These differences between groups did not emerge following masked CS presentations, suggesting that the between-group differences relied on explicit knowledge of the CS-US relationship during extinction. Critically however, immediate and delayed extinction groups also differed during extinction training; immediate extinction failed to produce successful extinction of CR, suggesting that recently acquired fear is more resistant to extinction and does not show recovery of fear. Although it is not clear how insufficient extinction affects the subsequent return of fear (Craske et al., 2008), the lack of complete extinction in the immediate group complicates the interpretation of the differences between immediate and delayed extinction.

4.3 STUDY III

Background and objectives

Recent progress in the field of fear learning has demonstrated that a single reminder exposure trial prior to extinction training can prevent the return of extinguished fear by disrupting the process of reconsolidation. Thus, extinction training initiated within but not outside of a critical reconsolidation time window was shown to block the return of fear, as first demonstrated in rodents (Monfils et al., 2009), and later extended to humans (Schiller et al., 2010) in a differential fear conditioning paradigm in which they used colored squares as CSs. However, recent failures to replicate these effects in humans (Kindt & Soeter, 2011; Soeter & Kindt, 2011) have raised the possibility that the discrepancies between studies are due to procedural differences. For instance, the effects reported by Schiller et al (2010) may not extend to fear-relevant stimuli. The aim of **Study III** was to further study the putative effects of disrupting reconsolidation. More specifically, we assessed whether extinction training initiated within the reconsolidation time window could abolish the return of fear using fear-relevant (experiment 1) or fear-irrelevant (experiment 2) CSs.

Methods

Experiment 1

Nineteen students at the Karolinska Institutet participated in the study that was run on three consecutive days conducted approximately 24 hr apart: acquisition (Day 1); reactivation and extinction (Day 2) and reinstatement and re-extinction (Day 3). On day 1, all participants were fear conditioned to three different pictures depicting fearful male faces from the Karolinska Directed Emotional Faces (Lundqvist et al., 1998), which served as conditioned stimuli (CSs). Each CS was presented for 6 s and a white fixation cross was shown on a black background during the ITIs (12-15 s). Two of the CSs (the to-be CS+r, and CS+nr) were coupled to an electric shock (the US), which was delivered to the lower ventral right arm on 6 of the 12 presentations of each CS+, whereas the third stimulus (the CS-) served as a control and was never coupled to a shock (12 non-reinforced presentations). On day 2, participants were exposed to a single unreinforced CS+ presentation (CS+r) in order to reactivate the acquisition memory. Which CS+ that was reactivated was counterbalanced across participants. Ten minutes after the reminder presentation, participants underwent extinction training consisting of non-reinforced presentations of all three CSs. Finally, on the third day of testing, the reinstatement and re-extinction session began with four unsignaled presentations of the US after which participants were given a 10-min break. Re-extinction followed immediately after the break and consisted of non-reinforced presentations of all CSs.

To assess conditioned fear responses, we measured the magnitude of the FPS. To elicit the startle reflex, startle probe (50 ms bursts of 95-dB[A] white noise with a near instantaneous rise time) were presented to the participants through headphones 4-5 s after stimulus onset on an equal number of trials of each CS and the ITIs (9 out of 12 trials during acquisition, 8 out of 12 during extinction, and 6 out of 9 during reinstatement testing). In addition, we measured SCR to each presentation of a CS throughout all experimental phases.

Experiment 2

In experiment 2, we enrolled 20 new participants and replaced the CSs used in experiment 1 with fear-irrelevant CSs, i.e. colored squares. Also, to increase comparability with the original study by Schiller et al (2010), and because it has been suggested that the presentation of startle probes might interfere with the measurement of SCR, we excluded startle probes in this second experiment and only measured SCRs. All other experimental parameters and procedures were identical to experiment 1.

Results and Conclusions

We found that a single reminder exposure trial prior to extinction training did not prevent the return of extinguished fear responding using either fear-relevant or fear-irrelevant CSs. The failure to replicate the study by Schiller et al (2010) has previously been discussed in terms of procedural differences between studies. Our results do not support the hypothesis that the failure to demonstrate that extinction training can disrupt reconsolidation is related to the fear-relevance of the CSs. Neither do our results support that the concurrent measurement of FPS, due to its intrinsically aversive nature, interferes with the measurement of SCR as we did not include auditory startle probes in experiment 2 and still found significant reinstatement of SCRs. Our findings, taken together with the fact that reconsolidation success is limited by several temporal boundary conditions of relevance to its potential to translate to the clinic, point to the need to further study the specific parameters that enable disruption of reconsolidation.

4.4 STUDY IV

Background and objectives

Much of what we learn about our environment is transferred from other individuals' behaviors through social forms of learning, such as observation. Past research has focused on how information about potential danger can be learned from observing another individual being subjected to threat. This has mainly been achieved by studying observational, or *vicarious*, learning of fear. However, little is known about how previously acquired fears can be attenuated through observation. The aim of **Study IV** was to develop an experimental paradigm to study the effects of vicarious safety learning on the decrement of learned fear during extinction training and its effects on the subsequent return of fear.

Method

A total of sixty-nine male students at Karolinska Institutet participated in the study in one single session. The experiment consisted of three different phases; fear acquisition, extinction and reinstatement testing. First, all participants were subjected to a fear conditioning protocol in which they watched a presentation of two different pictures depicting angry male faces from the Karolinska Directed Emotional Faces (Lundqvist et al., 1998), which served as conditioned stimuli (CSs). Each CS was presented for 6 s and a white fixation cross was shown on a black background during the ITIs (12-15 s). The presentation of one CS (the to-be CS+) was coupled to an electric shock (the US), which was delivered to the lower ventral right arm on six of the nine CS+ presentations, whereas the other stimulus (the CS-) served as a control and was never coupled to a shock (nine non-reinforced presentations). Which face that served as CS+ or CS- was counterbalanced between participants.

After this standard conditioning procedure, a Direct extinction and a Vicarious extinction group watched a movie involving six unreinforced presentations of the CS. For the Vicarious extinction group, a calmly looking learning model was depicted in the video as simultaneously watching the CS presentations. After the extinction session, all participants underwent a standard reinstatement procedure during which they received three unsignaled presentations of the US followed by six non-reinforced 6-s presentations of each of the two CSs. Also, to assess if the effects of vicarious extinction could be ascribed to the model's experience of non-reinforced CS presentations and not merely the presence of the learning model, we added a third group of subjects that also underwent all phases of the experiment. This *Vicarious reinforcement* group only differed from the Vicarious extinction group in that during extinction, the learning model received four shocks coupled with the presentations of the previously reinforced CS+. To assess conditioned fear responses, we measured SCR to each presentation of a CS throughout all experimental phases.

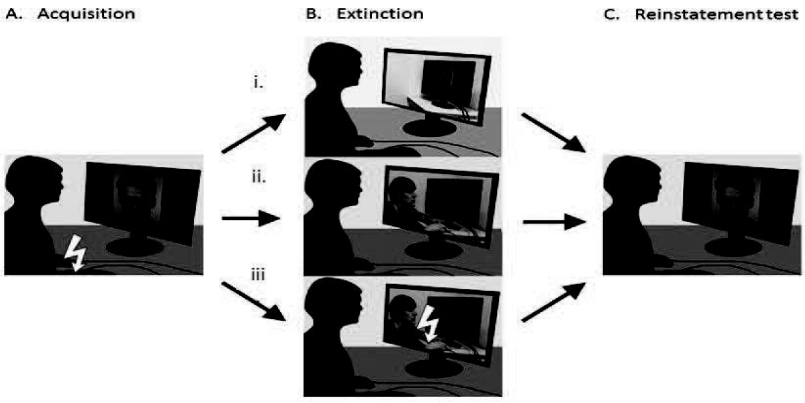

Figure 1. Experimental procedure. A. All participants underwent a standard fear conditioning protocol during which the presentation of one of two CSs was followed by a shock. B. During extinction, participants were divided into three different groups: i, Direct extinction ii,Vicarious extinction, and iii, Vicarious reinforcement. C. Finally, all participants underwent a standard reinstatement test during which they received three unsignaled shocks followed by the presentation of the CSs. See the main text for details.

Results and Conclusions

As compared to direct extinction, vicarious extinction promoted better extinction and blocked the return of fear during a subsequent reinstatement test. Also, by adding the Vicarious reinforcement group, we confirmed that these effects could be attributable to the model's experience of non-reinforced CS presentations and not simply to the presence of the learning model per se. This was achieved by demonstrating that the Vicarious extinction group also showed less reinstatement as compared to the Vicarious reinforcement group. Taken together, vicarious extinction efficiently reduced the expression of previously learned fear during both extinction training and during a subsequent reinstatement test. These findings may have important implications by integrating social and emotional learning processes.

5 GENERAL DISCUSSION

The overall aims of the studies presented in this thesis were to study the contribution of the temporal factors that govern the decrease of conditioned fear responding during extinction and to study different approaches to preventing the return of learned fear. Previous attempts to understand the processes underlying extinction have mainly focused on the contribution of associative and non-associative mechanisms by addressing whether extinction represents new learning of a CS–noUS association, an unlearning or erasure of the original CS-US association, a habituation-like process, or a combination of these mechanisms. The idea that extinction reflects learning of a new association is not new but has been extended by more recent accounts that have emphasized a critical role of context in determining the expression of extinction learning (see Bouton, 1993). The support for this so-called "new learning" account of extinction relies heavily on post-extinction phenomena during which extinguished fears reappear. In fact, the recovery of conditioned fear expression after extinction is a gold standard to conclude that extinction is not unlearning or erasure of the original memory trace, but entails the formation of a new CS-no US association.

In this section, I will start by discussing the temporal properties that critically drive extinction by discussing the results from **Study I**. Then I will, by discussing the results from **Study II-IV**, gradually turn to a more general discussion about how these studies collectively contribute to understanding how the return of extinguished fears can be prevented.

What causes extinction?

Traditionally, attempts at understanding the decrement in CR during extinction have focused on the associative nature of the extinction phenomena. A related strategy has sought to understand the decrement in CRs as a function of core temporal properties that are inherent in the extinction procedure. Thus, operationally, the decrement in CR results from repeated non-reinforced presentations of the CS. However, as the extinction procedure involves both an accumulating number of presentations of the CS, as well as an accumulating exposure time to the CS, it remains unclear which of these temporal variables that critically determines extinction. To address this fundamental issue, four different groups of participants were fear conditioned with a fixed CS-US interval that was followed by an extinction series in which the number of non-reinforced CS presentations and the total duration of non-reinforced exposure were independently manipulated.

The most conclusive finding from **Study I** was that an equal amount of exposure time did not result in equal extinction. This was evident when considering that, (a) at least initially, given the same amount of non-reinforced exposure time, the group that received the fewest number of trials showed least extinction as measured by FPS and

(b) given an equal number of non-reinforced trials, we observed facilitated extinction in the group that received the least cumulated non-reinforced exposure. In fact, extinction progressed most readily with many CS trials with a duration that was shorter than the original acquisition duration. Such an effect was not observed when lengthening the CS duration from acquisition to extinction, arguing against a generalization decrement hypothesis. This is noteworthy given that previous data from non-human animals, using similar designs but in different conditioning preparations, have generally shown that the degree of dissimilarity between the acquisition CS and the extinction CS duration critically affects response decrement during both extinction (Haselgrove & Pierce, 2003) and extinction re-test (Plendl & Wotjak, 2010; Drew et al., 2004).

According to componential trace models, the lack of non-reinforced presentations of the trained CS duration during extinction training is predicted to produce larger CRs when subsequently tested with the acquisition CS duration. Such an effect was not obvious in our data because the differences during extinction training did not predict the recovery of CR during a subsequent reinstatement test. This lack of transfer from extinction training performance to a subsequent extinction test is in line with previous work in both non-human animals (Drew et al., 2004; Haselgrove & Pierce, 2003; Plendl & Wotjack, 2010) and in humans (Prenoveau, 2012). These findings support previous proposals that extinction performance is not equivalent to extinction learning (Craske et al. 2008; but see Foa and Kozak, 1986) and may have important clinical implications given that extinction and recovery of fear represent the experimental analogue of therapeutic exposure and follow-up.

There is however one puzzling aspect of the results from **Study I**. Previous research has shown that CR are timed in relation to when the US is expected when training occurs with a fixed CS-US interval (e.g. Drew, Zupan, Cooke, Couvillon, & Balsam, 2005), i.e. the CS-US interval during acquisition is learned. In the absence of multiple indices of CR during a trial, the design from **Study I** does not allow us to address how the learned temporal relation between the CS and the US was affected by changing the CS duration from acquisition to extinction. Basically, there are (at least) three principal ways startle probe timing could have been assigned.

One option would have been to startle all groups at the same time point throughout the experiment, i.e. at 2s after CS onset. Then all groups would have had the same startle time point relative to CS onset but the groups would have varied in elapsed time from startle onset to US onset. Thus, if CS onset is the predictive cue for the US, then anchoring the startle probe onset relative to CS onset seems reasonable. However, if, as has been shown previously (Grillon, Ameli, Merikangas, Woods, & Davis, 1993), startle potentiation increases with temporal proximity to the US, then perhaps anchoring the startle probe time in relation to CS offset would be appropriate. This approach however inevitably allows the time from CS onset to the startle probe to vary.

Our rationale was to hold constant the timing of the startle probe relative to the total duration of the CS. This way, the elapsed time since CS onset, as well as the time to US

onset was scaled proportionally (halved or doubled) in the groups relative to the original training duration (6s). Nevertheless, this approach does not exclude alternative explanations of the data. To exclude that the startle probe timing confounded our extinction results, we analyzed and plotted the data from the first extinction trial for all groups separately and confirmed that the groups did not differ at the outset of extinction training, which would be expected if probe timing per se was sufficient to explain the differences between the groups. Also, equating groups based on exposure time did not reveal differences between the group receiving the shorter CS trial duration (3s) and the group receiving the acquisition CS trial duration (6s) even after 72 s of exposure, ruling out that the earlier probes in the group receiving the shorter CS trial durations were generally more sensitive to extinction effects.

Another interesting feature of **Study I** is that increasing the number of trials and exposure time did not affect differential fear responding when keeping CS duration constant, thus suggesting habituation of the startle response to both CSs at a similar rate. This is noteworthy for several reasons. First, we can rule out that the lack of complete differential responding is not merely a function of an insufficient number of extinction trials. Indeed, recent data from our lab (Golkar et al., 2012) demonstrate that 12 CS exposure trials is sufficient to extinguish FPS responses to fear-irrelevant (colored squares), but not fear-relevant (angry faces) stimuli in a within-subject design. This suggests that the lack of extinguished conditioned fear responses during immediate extinction might be related to the fear-relevant properties of the CSs. Second, although we observed a significant increase in the CR to the CS+ after reinstatement testing, we did not observe a differential increase in conditioned fear responding in any of the extinction groups in **Study I**, which could reflect the involvement of non-associative processes such as sensitization to the reinstatement shocks and/or generalization of fear to the context. I will return to these issues in relation to the results from **Study II**.

Can extinction erase fears?

Recently, considerable attention has been turned to studying the potential determinants of the mechanisms that are engaged during extinction by focusing on the temporal relationship between fear acquisition and extinction on the one hand, and fear reactivation and extinction, on the other. This interest is largely motivated by the fact that the findings may have important clinical implications by identifying the temporal intervals during which behavioral interventions may permanently prevent the recovery of learned fears.

In **Study II**, based on previous findings in rodents suggesting that extinction initiated immediately after conditioning reflects memory erasure (Myers et al., 2006), we tested the influence of varying the temporal delay between acquisition and extinction on the return of fear in a differential fear conditioning paradigm in humans. Our main findings were generally in line with Myers and colleagues (Myers et al., 2006); immediate

extinction, in contrast to delayed extinction, did not result in a significant return of fear as measured by reinstatement of FPS. Moreover, we manipulated contingency awareness during extinction by including both masked and non-masked extinction conditions to confirm that the apparent differences between immediate and delayed extinction were specifically related to conditions that allowed for explicit CS-US contingency learning.

Although these results suggest that immediate extinction blocked the reinstatement of FPS, other studies that have explicitly manipulated acquisition-to-extinction timing, albeit using different experimental conditions and other measures of CR, have not demonstrated this effect (Archbold, Bouton, & Nader, 2010; Huff, Hernandez, Blanding, & LaBar, 2009; Woods & Bouton, 2008) (but see Norrholm et al., 2008 for partly overlapping findings using FPS in humans). Given these inconsistencies, perhaps the relevant question is not a categorical one of whether immediate extinction causes an erasure of fear, but rather under which conditions it suppresses conditioned responding and whether these conditions can be reliably reproduced.

In a related line of studies, Maren and Chang (2006) reported that immediate extinction caused a short-term suppression of conditioned freezing in rats that fully recovered the following day (Maren & Chang, 2006). This lack of long-term fear suppression following immediate extinction has been termed the "immediate extinction deficit". This deficit lasted even when extinction training was initiated up to 6 hr after fear acquisition (Chang & Maren, 2009), suggesting that there is a time window following fear acquisition during which fear memory is resistant to the effects of extinction. Interestingly, the authors reported in two separate studies (Chang & Maren, 2009; Maren & Chang, 2006) that the suppression of fear during immediate, but not delayed, extinction training was similar in rats exposed to non-reinforced CS trials during extinction as in a group of rats that only received context exposure in the absence of non-reinforced CS trials (no-extinction control group). Moreover, in contrast to the effects of delayed extinction, the suppression of fear following immediate extinction was insensitive to contextual manipulations, which is a hallmark of extinction learning. Collectively, these observations led Chang and Maren (2009) to suggest that the reduction of responding that occurs with non-reinforced presentations of the CS shortly after fear acquisition might reflect a context-independent habituation-like mechanism. Interestingly, subsequent work has showed that immediate extinction does not recruit mPFC circuits that are implicated in successful extinction learning (Chang, Berke, & Maren, 2010; Kim, Jo, Kim, Kim, & Choi, 2010). This supports that immediate extinction might, at least partly, recruit different processes than those during delayed extinction.

It is tempting to analyze the effects from **Study II** in terms of the habituation-proposal by Chang & Maren (2009). Supporting their proposal, extinction and habituation have been shown to share several fundamental properties (McSweeney & Swindell, 2002), and it is not clear to what extent habituation-like mechanisms contribute to extinction. Although the results from **Study II** and that of Maren & Chang (2006) are seemingly

different in that Maren & Chang (2006) emphasized that immediate extinction did not eliminate long-term CR, we did not assess the long-term effects of immediate and delayed extinction using a delayed extinction-to-test interval (see also Johnson, Escobar, & Kimble, 2010). In the short-term however, our results overlap in that both studies showed that immediate extinction resulted in a suppression of conditioned fear that was not evident following delayed extinction. Given the controversies in the short and long-term effects of immediate extinction, it remains unclear whether initiating extinction training shortly after fear acquisition can interfere with the consolidation of fear memory to reduce the subsequent expression of fear, and to what extent the effects of immediate extinction may be mediated by habituation-like mechanisms.

There is an inevitable complication in interpreting the recovery data in **Study II** as the immediate and delayed groups also differed in the degree of within-session extinction, i.e. the inability to extinguish differential CR during extinction training was restricted to the immediate extinction group. Collectively however, the results from **Study I** and **Study II** extends previous findings of resistance to extinction with fear-relevant CSs (Öhman & Mineka, 2001), by suggesting that this resistance to extinction is insensitive to increasing the number of CS trials and exposure time to the CS (**Study I**), is restricted to immediate extinction (**Study II**), and does not result in differential responding during a short-interval reinstatement test conducted after immediate extinction (**Study I and II**).

From a clinical perspective, the restricted time window after fear acquisition during which memory is susceptible to disruption undoubtedly restrains the applicability of interventions. However, similar temporal time windows have been observed shortly after retrieval of fear memory (Nader & Hardt, 2009; Sara, 2000) during which previously consolidated memories can be reactivated and again rendered sensitive to disruption. This reconsolidation process has received considerable attention during the last decades much owing to a series of studies demonstrating that manipulations interfering with fear memory consolidation also disrupt fear memory when administered shortly after retrieval of that memory (Nader et al., 2000). Two influential lines of research have emerged from these demonstrations. First, beta-adrenergic receptor blockade can disrupt reconsolidation and prevent the subsequent expression of fear in both rodents (e.g. Debiec & LeDoux, 2004) and humans (e.g. Kindt, Soeter & Vervliet, 2009). Second, extinction training initiated within this critical reconsolidation time window has been reported to produce similar effects on the return of fear (e.g. Monfils et al., 2009; Schiller et al., 2010).

Given the tremendous clinical implications of preventing the expression of acquired fear memories with a behavioral intervention, coupled with the fact that these effects have proven hard to replicate using more clinically relevant stimuli, i.e. fear-relevant stimuli such as spiders, (Kindt & Soeter, 2011; Soeter & Kindt, 2011), we attempted to replicate and extend the findings from Schiller et al (2010). Therefore, in **Study III**, we assessed whether extinction training initiated within the reconsolidation time window could abolish the expression of fear during a subsequent recovery test using both fear-

relevant and fear-irrelevant stimuli in two separate experiments. Our main finding was that extinction training following reactivation of the fear memory did not prevent the recovery of fear, as measured by reinstatement of the FPS or SCR using either fear-relevant or fear-irrelevant stimuli. As such, these negative results are in line with previous replication failures (Kindt & Soeter, 2011; Soeter & Kindt, 2011) and contrast with more recent studies that have demonstrated similar effects to those of Schiller et al (2010) using SCR (Oyarzun et al., 2012) and BOLD responses in the amygdala (Ågren et al., 2012) as measures of CR.

The inconsistencies between previous reports have been speculated to reflect procedural differences, (Kindt & Soeter, 2011; Oyarzun et al., 2012), such as differences in memory strength related to differences in acquisition reinforcement rates or the fear-relevant properties of the CSs, and the use of concurrent indices of CR that presumably cause interference between measurements. Our results gives limited support for these possible explanations, but nevertheless leaves the fundamental question unresolved; what conditions do allow for erasure of fear memory?

Indeed, reconsolidation is bounded by several conditions. Thus, previous work has shown that interference with reconsolidation is temporally graded, such that recent memories are more sensitive to disruption than more remote memories (Frankland et al., 2006; Suzuki, Josselyn, Frankland, Masushige, Silva, & Kida, 2004) and that the temporal dynamics of reconsolidation are dependent on the strength of the acquired memory (Suzuki, Josselyn, Frankland, Masushige, Silva, & Kida, 2004; Wang et al., 2009). Although it is still unclear exactly how these boundary conditions explain the failure to disrupt reconsolidation in **Study III**, replication failures like ours and those of Kindt and Soeter (Kindt & Soeter, 2011; Soeter & Kindt, 2011) raise the question of whether the reconsolidation effects demonstrated by Schiller et al (2010) are stable enough to be translated into the highly complex situations in which fears are acquired and expressed. Notably however, using a related strategy, Soeter & Kindt (2011) have shown that the same parameters that enable disruption of fear expression using beta-adrenergic receptor blockade during reconsolidation (as indexed by FPS) did not prevent the return of fear with only extinction training initiated subsequent to retrieval. Importantly, in a series of experiments, Kindt and collegues have demonstrated the efficacy of administering the beta-adrenergic receptor antagonist propranolol either prior (Kindt, Vervliet & Soeter, 2009; Soeter & Kindt, 2010; Soeter & Kindt, 2011; Kindt & Soeter, 2011) or post-reactivation (Soeter & Kindt, 2012a; Soeter & Kindt, 2012b) on the return of fear. Importantly, the effects were restricted to FPS responses, whereas SCRs and CS-US contingency ratings remained unaffected by this manipulation, suggesting that successful reconsolidation interference may require different reactivation conditions. As argued by Soeter & Kindt (2012a), this view is in line with a functional account of reconsolidation, emphasizing that reconsolidation is an integral part of memory modification and storage; its functional role is to update memories to maintain their relevance (Lee et al., 2009). Moreover, the findings reported by Soeter & Kindt are further strengthened by the extension of the efficacy of disrupting reconsolidation with propranolol on fears acquired through verbal

instructions (Soeter & Kindt, 2012a) and a recent demonstration of these effects on the subjective levels of anxiety, which is of particular clinical relevance. It remains to be shown if future research on reconsolidation using behavioral interventions will unravel the conditions that reliably reproduce the fear attenuating effects of such interventions.

As an alternative approach to preventing the return of fear, in **Study IV**, we capitalized on the fact that much of what we learn about the environment, such as information about what should be avoided and approached, comes through social forms of learning such as through instruction from or observation of other individuals. Indeed, learning from others' experiences is often less risky in comparison to self-experienced trial and error and perhaps owing to the cost-benefits of such learning, some forms of social learning have been well conserved across social animals (Jeon et al., 2010; Mineka & Cook, 1993; Olsson & Phelps, 2007).

In **Study IV**, we developed an experimental paradigm to study the attenuation of learned fears through social observation. This was accomplished by assessing how directly experienced fear learning could be extinguished by observing another individual being exposed to unreinforced presentations of the fear-eliciting stimulus (vicarious extinction) as compared to direct non-reinforced exposure (direct extinction) or compared to observing another individual being exposed to reinforced exposures (vicarious reinforcement). The main results from **Study IV** were that, compared to both the Direct extinction group and the Vicarious reinforcement group, vicarious extinction efficiently reduced CR during extinction and blocked the subsequent return of fear as measured by differential SCR to the CS+ vs. the CS- during a subsequent reinstatement test. As such, our results may have important implications for clinical practice by integrating social and emotional learning processes.

However, it is not clear from our findings *how* vicarious extinction exerts its extinction-facilitating effects. By adding the Vicarious reinforcement group we can rule out that this process was driven simply by the presence of the learning model. Rather, it seems to be driven, at least to some extent, by the content of the learning model's experience. For instance, the presence of the shocks during vicarious reinforcement might have additionally strengthened the previously learned CS-US association through observational learning mechanisms, but this does not explain why extinction was more efficient in the Vicarious extinction group compared to the Direct extinction group. Drawing upon the known mechanisms of vicarious fear learning (Olsson & Phelps, 2007; Mineka & Cook, 1993), the results from **Study IV** indicate that watching the calm learning model during extinction imbued the CS+ with a safety value by recruiting additional processes than those shared with direct extinction. The extent to which such processes depend on higher cognitive and social inference mechanisms remains poorly understood, but is highlighted by the fact that the differences between the Vicarious extinction group and the Direct extinction group emerged in spite of the fact that they received an equal amount of non-reinforced CS exposure. However, it is not clear how to separate such cognitive inference processes, i.e. inferring that a situation is dangerous or safe to oneself from the behavior of another organism, from a

more simple associative process given that even simple CS-US learning entails learning about the causal relationships between events (Rescorla, 1988; Dickinson, 1980). In fact, as argued by Mineka and Cook (1993) in relation to observational fear learning, if the organism is making something very akin to cognitive causal inference during conditioning, (i.e. the organism is attempting to create a causal structure of its environment), then the differences between these processes may only be visible at a superficial level of analysis, which does not exclude that they are mediated by essentially the same mechanisms. Applying the same logic to interpret the results from **Study IV** it is reasonable to assume that vicarious and direct extinction rely on the same underlying mechanisms, but that the presence of the learning model adds additional information that can be used to infer the safety value of the CS.

Critically however, in order to live up to their clinical potential, the findings from **Study IV** require both replication using a delayed extinction protocol to assess the effects of vicarious extinction on consolidated fear memories, and an assessment of its efficacy using a delayed extinction-test interval to assess the long-term effects of such treatment. Moreover, the documented dissociations between response systems, such as that between FPS and SCR (e.g. Soeter & Kindt, 2010; Weike et al., 2005), highlight the importance of using multiple indices of CR in future studies addressing the effects of vicarious extinction.

Perhaps most relevant in terms of the erasure mechanisms of extinction is the fact that the lack of CR after vicarious extinction resembles those previously described in relation to disruption of consolidation (e.g. Myers et al., 2006) and reconsolidation (e.g. Kindt et al., 2009; Schiller et al., 2010). That is, the absence of CR during a subsequent recovery test, if taken at face value, indicates that the fear memory was erased during vicarious extinction. This interpretation however, requires some caution. First, in contrast to the previous conditions during which fear memories have been suggested to be erased (i.e. immediate extinction or post-reactivation), there is yet no mechanistic explanation for how vicarious extinction would exert such memory erasing effects, which limits its theoretical appeal. Second, an overlapping pattern of observed behavior (i.e. absence of CR) does not necessarily imply overlapping mechanisms. In fact, the absence of CR during recovery tests is in itself not sufficient to index erasure as there are several alternative accounts to explain these phenomena (Delamater, 2004).

As argued by Lattal and Stafford (2008), demonstrating that behavior fails to show spontaneous recovery, renewal, or reinstatement after extinction is not sufficient to index that a manipulation either erased the original memory or enhanced the extinction memory. For instance, there are several possible explanations to the absence of complete recovery. Whereas it could mean that part of the memory was erased, it could also mean that the particular experimental conditions lacked sensitivity to detect the recovery of fear. Also, keeping in mind that learning and memory are constructs that are inferred from behavior, the absence of behavior does not necessarily mean absence of memory (e.g. Stout & Miller, 2007). Perhaps then, more convincing evidence regarding the erasure of memories will come from other levels of analysis. As an

example, it has previously been shown that immediate, but not delayed extinction can reverse some of the molecular substrates of learning-related synaptic plasticity that is induced by fear conditioning (Mao et al., 2006), thereby providing a putative neurobiological explanation of erasure. Although it is tempting to conclude that a mechanistic account of fear erasure will come from molecular studies, it is important to note that without a complete understanding of the conditions that cause memory formation, we are unlikely to be able to firmly establish that a fear memory has been retroactively erased by extinction. Obviously, a fuller understanding of the mechanisms that govern fear memory formation and erasure will emerge from studies across levels of analyses that focus on assessing how different manipulations affect extinction learning and whether these effects are persistent across time and context.

Taken together, the results from the studies in this thesis, together with the wealth of data that have accumulated across the last century, highlight that extinction, even at a behavioral level of analysis, represents a highly complex phenomenon that most probably is determined by multiple factors.

6 FUTURE DIRECTIONS

Understanding learning is fundamental to understanding the behavior of an organism as it initiates the process whereby our experiences of the environment can translate into new behaviors and change or modify those that have already been established. Extinction learning, in particular, is often presented as a leading paradigm to study how humans and other organisms can learn to flexibly adapt behavior to changing environmental contingencies. In this respect, there are two major lines of research that await more attention.

First, the field of fear learning and extinction has put considerable more effort into understanding the mechanisms and processes involved in the *reactions* as compared to the *instrumental actions* involved in fear learning. A shift of attention toward actions may be of particular value for increasing the translational validity of Pavlovian fear conditioning models. For instance, focusing on action tendencies that capture the conflict between approach-avoidance behaviors would incorporate one of the cardinal behavioral features of different anxiety disorders. Indeed, avoidance behaviors are thought to be critical in maintaining anxiety and are the target of the majority of behavioral treatments of anxiety (Barlow, 2002).

Second, the procedures commonly used to assess Pavlovian fear learning in humans do not allow separation of predictive fear learning from simple contiguity learning. Thus, given that Pavlovian fear learning is sensitive to both the contiguous and predictive relationship between stimuli, distinguishing between these different relations requires somewhat different approaches (see McNally & Westbrook, 2006 for a discussion). Fortunately, there are several models developed in non-human animals that serve these purposes. Such cue competition models appreciate that the relation between cues in the environment is often quite ambiguous and that different cues can compete for threat-relevance. Such models could also prove informative in clinical contexts because they incorporate a level of complexity that more resembles real-life learning situations (see Beckers, Krypotos, Boddez, Effting, & Kindt, 2013 for a related discussion).

Indeed, modeling the complexity of real-life situations poses a challenge to future experimental research on fear learning and extinction. The challenge involves continuing to capitalize on the benefits of focusing on the commonalities in evolved mechanisms between humans and non-human animals but also develop models that can appreciate the differences between the computations allowed by the human brain and the brains of other species, and incorporate the influence of the complex social environments with which humans interact. Appreciating these notions are obviously relevant in explaining human behavior, but perhaps more so when explaining dysfunctions in human behavior.

7 ACKNOWLEDGEMENTS

I wish to express my gratitude to my main supervisor Arne Öhman and my co-supervisor Martin Ingvar.

Arne, thank you for sharing your dedication to science with me. I began my PhD work inspired by what you had accomplished and I will continue to pursue research, still inspired by you.

Martin, thank you for invaluable support with the fMRI studies that turned out to be beyond the scope of this thesis.

In addition I would like to thank:

My colleague Andreas Olsson, for your contagious curiosity in science, your overwhelming optimism, and your friendship.

My former colleagues at the Department of Psychology at Stockholm University. Especially Stefan Wiens for offering me excellent scientific training, and Nathalie Peira for your genuine support throughout.

My former and present colleagues at the Section of Psychology and "old" MR-centrum. Especially those that I've worked closest with; Tina Lonsdorf, Jonathan Berrebi, Kara Lindstrom, Ida Selbing, and Tanaz Molapour.

The four prodigies (in order of appearance at the lab): Rasmus Berggren, Daniel Fürth, Martin Bellander, and Björn Lindström; for sharing your brilliant minds.

And finally Gustaf;

for everything beyond temporal properties of fear extinction.

8 REFERENCES

Ågren, T., Engman, J., Frick, A., Björkstrand, J., Larsson, E.-M., Furmark, T., & Fredrikson, M. (2012). Disruption of reconsolidation erases a fear memory trace in the human amygdala. *Science, 337*.

Alberini, C. M. (2005). Mechanisms of memory stabilization: Are consolidation and reconsolidation similar or distinct processes? *Trends in Neurosciences, 28*(1), 51-56.

Archbold, G. E. B., Bouton, M. E., & Nader, K. (2010). Evidence for the persistence of contextual fear memories following immediate extinction. *European Journal of Neuroscience, 31*(7), 1303-1311.

Askew, C., & Field, A. P. (2008). The vicarious learning pathway to fear 40 years on. *Clinical Psychology Review, 28*(7), 1249-1265.

Bandura, A. (1977). *Social learning theory*. New York.

Bandura, A., Grusec, J. E., & Menlove, F. L. (1967). Vicarious extinction of avoidance behavior. *Journal of Personality and Social Psychology, 5*(1), 16-&.

Bandura, A., & Rosenthal, T. L. (1966). Vicarious classical conditioning as a function of arousal level. *Journal of Personality and Social Psychology, 3*(1), 54-&.

Bandura, A., Ross, S. A., & Ross, D. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology, 63*(3), 575-&.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science, 269*(5227), 1115-1118.

Beckers, T., Krypotos, A. M., Boddez, Y., Effting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biological Psychology, 92*(1), 90–96.

Berger, S. M. (1962). Conditioning through vicarious instigation. *Psychological Review, 69*(5), 450-466.

Blair, H. T., Schafe, G. E., Bauer, E. P., Rodrigues, S. M., & LeDoux, J. E. (2001). Synaptic plasticity in the lateral amygdala: A cellular hypothesis of fear conditioning. *Learning & Memory, 8*(5), 229-242.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of pavlovian learning. *Psychological Bulletin, 114*(1), 80-99.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*(10), 976–986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*, 485-494.

Bouton, M. E., & Bolles, R. C. (1979a). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10*(4), 445-466.

Bouton, M. E., & Bolles, R. C. (1979b). Role of conditioned contextual stimuli in reinstatement of extinguished fear. *Journal of Experimental Psychology-Animal Behavior Processes, 5*(4), 368-378.

Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear - test for the associative value of the context. *Journal of Experimental Psychology-Animal Behavior Processes, 9*(3), 248-265.

Bouton, M. E., Westbrook, R. F., Corcoran, K. A., & Maren, S. (2006). Contextual and temporal modulation of extinction: Behavioral and biological mechanisms. *Biological Psychiatry, 60*(4), 352-360.

Brandon, S. E., Vogel, E. H., & Wagner, A. R. (2002). A neuroscientist's guide to classical conditioning. In J. W. Moore (Ed.), *Computational theories of classical conditioning* (pp. 232–310). New York: Springer-Verlag

Brooks, D. C., & Bouton, M. E. (1993). A retrieval cue for extinction attenuates spontaneous-recovery. *Journal of Experimental Psychology-Animal Behavior Processes, 19*(1), 77-89.

Brunet, A., Orr, S. P., Tremblay, J., Robertson, K., Nader, K., & Pitman, R. K. (2008). Effect of post-retrieval propranolol on psychophysiologic responding during subsequent script-driven traumatic imagery in post-traumatic stress disorder. *Journal of Psychiatric Research, 42*(6), 503-506.

Buchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain systems mediating aversive conditioning: An event-related fmri study. *Neuron, 20*(5), 947-957.

Burgos-Robles, A., Vidal-Gonzalez, I., Santini, E., & Quirk, G. J. (2007). Consolidation of fear extinction requires nmda receptor-dependent bursting in the ventromedial prefrontal cortex. *Neuron, 53*(6), 871-880.

Cain, C. K., Blouin, A. M., & Barad, M. (2004). Adrenergic transmission facilitates extinction of conditional fear in mice. *Learning & Memory, 11*(2), 179-187.

Chang, C. H., Berke, J. D., & Maren, S. (2010). Single-unit activity in the medial prefrontal cortex during immediate and delayed extinction of fear in rats. *PLoS ONE, 5*(8).

Chang, C. H., & Maren, S. (2009). Early extinction after fear conditioning yields a context-independent and short-term suppression of conditional freezing in rats. *Learning & Memory, 16*(1), 62-68.

Cheng, D. T., Knight, D. C., Smith, C. N., & Helmstetter, F. J. (2006). Human amygdala activity during the expression of fear responses. *Behavioral Neuroscience, 120*(6), 1187-1195.

Chhatwal, J. P., Davis, M., Maguschak, K. A., & Ressler, K. J. (2005). Enhancing cannabinoid neurotransmission augments the extinction of conditioned fear. *Neuropsychopharmacology, 30*(3), 516-524.

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy, 46*(1), 5-27.

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience, 15*, 353-375.

Davis, M. (1997). The neurophysiological basis of acoustic startle modulation: Research on fear motivation and sensory gating. In P. J. Lang, R. F. Simons & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 69–96). Mahwah, N.J.: Lawrence Erlbaum Associates.

Davis, M., Walker, D. L., & Myers, K. M. (2003). Role of the amygdala in fear extinction measured with potentiated startle. *Annals of the New York Academy of Sciences, 985*, 218-232.

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry, 6*(1), 13-34.

Debiec, J., Bush, D. E. A., & LeDoux, J. E. (2011). Noradrenergic enhancement of reconsolidation in the amygdala impairs extinction of conditioned fear in rats-a possible mechanism for the persistence of traumatic memories in ptsd. *Depression and Anxiety, 28*(3), 186-193.

Debiec, J., & LeDoux, J. E. (2004). Disruption of reconsolidation but not consolidation of auditory fear conditioning by noradrenergic blockade in the amygdala. *Neuroscience, 129*(2), 267-272.

Delamater, A. R. (2004). Experimental extinction in pavlovian conditioning: Behavioural and neuroscience perspectives. *Quarterly Journal of Experimental*

Psychology Section B-Comparative and Physiological Psychology, 57(2), 97-132.

Dickinson, A. (1980). *Contemporary learning theory*. Cambridge, UK: Cambridge University Press.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*(5), 549-554.

Drew, M. R., Yang, C., Ohyama, T., & Balsam, P. D. (2004). Temporal specificity of extinction in autoshaping. *Journal of Experimental Psychology-Animal Behavior Processes, 30*(3), 163-176.

Drew, M. R., Zupan, B., Cooke, A., Couvillon, P. A., & Balsam, P. D. (2005). Temporal control of conditioned responding in goldfish. *Journal of Experimental Psychology-Animal Behavior Processes, 31*(1), 31-39.

Dudai, Y., & Eisenberg, M. (2004). Rites of passage of the engram: Reconsolidation and the lingering consolidation hypothesis. *Neuron, 44*(1), 93-100.

Dunsmoor, J. E., Bandettini, P. A., & Knight, D. C. (2008). Neural correlates of unconditioned response diminution during pavlovian conditioning. *Neuroimage, 40*(2), 811-817.

Duvarci, S., & Nader, K. (2004). Characterization of fear memory reconsolidation. *Journal of Neuroscience, 24*(42), 9269-9275.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*(9), 345-352.

Esteves, F., Dimberg, U., & Öhman, A. (1994). Automatically elicited fear - conditioned skin-conductance responses to masked facial expressions *Cognition & Emotion, 8*(5), 393-413.

Falls, W. A., Miserendino, M. J., & Davis, M. (1992). Extinction of fear-potentiated startle: Blockade by infusion of an nmda antagonist into the amygdala. *Journal of Neuroscience, 12*(3), 854-863.

Fendt, M., & Fanselow, M. S. (1999). The neuroanatomical and neurochemical basis of conditioned fear. *Neuroscience & Biobehavioral Reviews, 23*(5), 743-760.

Foa, E. B. (2000). Psychosocial treatment of posttraumatic stress disorder. *Journal of Clinical Psychiatry, 61*, 43-51.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear - exposure to corrective information. *Psychological Bulletin, 99*(1), 20-35.

Frankland, P. W., Ding, H. K., Takahashi, E., Suzuki, A., Kida, S., & Silva, A. J. (2006). Stability of recent and remote contextual fear memory. *Learning & Memory, 13*(4), 451-457.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review, 107*(2), 289-344.

Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: Interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience, 115*(4), 1261-1279.

Gilroy, L. J., Kirkby, K. C., Daniels, B. A., Menzies, R. G., & Montgomery, I. M. (2001). Controlled comparison of computer-aided vicarious exposure versus live exposure in the treatment of spider phobia. *Behavior Therapy, 31*(4), 733-744.

Globisch, J., Hamm, A. O., Esteves, F., & Öhman, A. (1999). Fear appears fast: Temporal course of startle reflex potentiation in animal fearful subjects. *Psychophysiology, 36*(1), 66-75.

Golkar, A., Fürth, D., Öhman, A., & Olsson, A. (2012). Extinction of conditioned fear to fear-relevant and fear-irrelevant stimuli. Manuscript in preparation.

Gottfried, J. A., & Dolan, R. J. (2004). Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nature Neuroscience, 7*(10), 1144-1152.

Grace, A. A., & Rosenkranz, J. A. (2002). Regulation of conditioned responses of basolateral amygdala neurons. *Physiology & Behavior, 77*(4-5), 489-493.

Grillon, C., Ameli, R., Merikangas, K., Woods, S. W., & Davis, M. (1993). Measuring the time-course of anticipatory anxiety using the fear-potentiated startle reflex. *Psychophysiology, 30*(4), 340-346.

Grossberg, S., & Schmajuk, N. A. (1991). Neural dynamics of adaptive timing and temporal discrimination during associative learning. In S. Grossberg (Ed.), *Pattern recognition by self-organizing neural networks* (pp.637–674). Cambridge: The MIT Press.

Guastella, A. J., Dadds, M. R., Lovibond, P. F., Mitchell, P., & Richardson, R. (2007). A randomized controlled trial of the effect of d-cycloserine on exposure therapy for spider fear. *Journal of Psychiatric Research, 41*(6), 466-471.

Guastella, A. J., Lovibond, P. F., Dadds, M. R., Mitchell, P., & Richardson, R. (2007). A randomized controlled trial of the effect of d-cycloserine on extinction and fear conditioning in humans. *Behaviour Research & Therapy, 45*(4), 663-672.

Hamm, A. O., & Weike, A. I. (2005). The neuropsychology of fear learning and fear regulation. *International Journal of Psychophysiology, 57*(1), 5-14.

Haselgrove, M., & Pearce, J. M. (2003). Facilitation of extinction by an increase or a decrease in trial duration. *Journal of Experimental Psychology-Animal Behavior Processes, 29*(2), 153-166.

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biological Psychiatry, 60*(4), 361-368.

Hermans, D., Dirikx, T., Vansteenwegenin, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*(4), 533-551.

Herry, C., Ferraguti, F., Singewald, N., Letzkus, J. J., Ehrlich, I., & Luthi, A. (2010). Neuronal circuits of fear extinction. *European Journal of Neuroscience, 31*(4), 599-612.

Hill, J. H., Liebert, R. M., & Mott, D. E. W. (1968). Vicarious extinction of avoidance behavior through films - an initial test. *Psychological Reports, 22*(1), 192-&.

Hobin, J. A., Goosens, K. A., & Maren, S. (2003). Context-dependent neuronal activity in the lateral amygdala represents fear memories after extinction. *Journal of Neuroscience, 23*(23), 8410-8416.

Hofmann, S. G., Pollack, M. H., Otto, M. W. (2006). Augmentation treatment of psychotherapy for anxiety disorders with d-cycloserine. *CNS Drug Reviews, 12*(3-4), 208-217.

Hooker, C. I., Verosky, S. C., Miyakawac, A., Knight, R. T., & D'Esposito, M. (2008). The influence of personality on neural mechanisms of observational fear and reward learning. *Neuropsychologia, 46*(11), 2709-2724.

Huff, N. C., Hernandez, J. A., Blanding, N. Q., & LaBar, K. S. (2009). Delayed extinction attenuates conditioned fear renewal and spontaneous recovery in humans. *Behavioral Neuroscience, 123*(4), 834-843.

Hygge, S., & Öhman, A. (1978). Modeling processes in acquisition of fears - vicarious electrodermal conditioning to fear-relevant stimuli. *Journal of Personality and Social Psychology, 36*(3), 271-279.

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S. Y., . . . Shin, H. S. (2010). Observational fear learning involves affective pain system and ca(v)1.2 ca2+ channels in ACC. *Nature Neuroscience, 13*(4), 482-U105.

Johnson, J. S., Escobar, M., & Kimble, W. L. (2010). Long-term maintenance of immediate or delayed extinction is determined by the extinction-test interval. *Learning & Memory, 17*(12), 639-644.

Kalisch, R., Korenfeld, E., Stephan, K. E., Weiskopf, N., Seymour, B., & Dolan, R. J. (2006). Context-dependent human extinction memory is mediated by a ventromedial prefrontal and hippocampal network. *Journal of Neuroscience, 26*(37), 9503-9511.

Kamprath, K., Marsicano, G., Tang, J. R., Monory, K., Bisogno, T., Di Marzo, V., . . . Wotjak, C. T. (2006). Cannabinoid cb1 receptor mediates fear extinction via habituation-like processes. *Journal of Neuroscience, 26*(25), 6677-6686.

Kamprath, K., & Wotjak, C. T. (2004). Nonassociative learning processes determine expression and extinction of conditioned fear in mice. *Learning & Memory, 11*(6), 770-786.

Kaplan, G. B., & Moore, K. A. (2011). The use of cognitive enhancers in animal models of fear extinction. *Pharmacology Biochemistry and Behavior, 99*(2), 217-228.

Kelly, M. M., & Forsyth, J. R. (2009). Associations between emotional avoidance, anxiety sensitivity, and reactions to an observational fear challenge procedure. *Behaviour Research and Therapy, 47*(4), 331-338.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport, 14*(18), 2317-2322.

Kim, S. C., Jo, Y. S., Kim, I. H., Kim, H., & Choi, J. S. (2010). Lack of medial prefrontal cortex activation underlies the immediate extinction deficit. *Journal of Neuroscience, 30*(3), 832-837.

Kindt, M., & Soeter, M. (2011). Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. Biological psychology.

Kindt, M., Soeter, M., & Vervliet, B. (2009). Beyond extinction: Erasing human fear responses and preventing the return of fear. *Nature Neuroscience, 12*(3), 256-258.

Knight, D. C., Nguyen, H. T., & Bandettini, P. A. (2005). The role of the human amygdala in the production of conditioned fear responses. *Neuroimage, 26*(4), 1193-1200.

Knight, D. C., Smith, C. N., Cheng, D. T., Stein, E. A., & Helmstetter, F. J. (2004). Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning. *Cognitive, Affective & Behavioral Neuroscience, 4*(3), 317-325.

Kushner, M. G., Kim, S. W., Donahue, C., Thuras, P., Adson, D., Kotlyar, M., . . . Foa, E. B. (2007). D-cycloserine augmented exposure therapy for obsessive-compulsive disorder.[see comment]. *Biological Psychiatry, 62*(8), 835-838.

Kwapis, J. L., Jarome, T. J., Schiff, J. C., & Helmstetter, F. J. (2011). Memory consolidation in both trace and delay fear conditioning is disrupted by intra-amygdala infusion of the protein synthesis inhibitor anisomycin. *Learning & Memory, 18*(11), 728-732.

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fmri study. *Neuron, 20*(5), 937-945.

LaBar, K. S., LeDoux, J. E., Spencer, D. D., & Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *Journal of Neuroscience, 15*(10), 6846-6855.

LaBar, K. S., & Phelps, E. A. (2005). Reinstatement of conditioned fear in humans is context dependent and impaired in amnesia. *Behavioral Neuroscience, 119*(3), 677-686.

Lang, P. J., Davis, M., & Öhman, A. (2000). Fear and anxiety: Animal models and human cognitive psychophysiology. *Journal of Affective Disorders, 61*(3), 137-159.

Lattal, K. M., & Stafford, J. M. (2008). What does it take to demonstrate memory erasure? Theoretical comment on norrholm et al. (2008). *Behavioral Neuroscience, 122*(5), 1186-1190.

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 years of consolidation - remembering muller and pilzecker. *Learning & Memory, 6*(2), 77-87.

Ledgerwood, L., Richardson, R., & Cranney, J. (2003). Effects of d-cycloserine on extinction of conditioned freezing. *Behavioral Neuroscience, 117*(2), 341-349.

Ledgerwood, L., Richardson, R., Cranney, J. (2005). D-cycloserine facilitates extinction of learned fear: Effects on reacquisition and generalized extinction. *Biological Psychiatry, 57*(8), 841-847.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*, 155-184.

Lee, J. L. C. (2009). Reconsolidation: Maintaining memory relevance. *Trends in Neurosciences, 32*(8), 413-420.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *KDEF: The Karolinska Directed Emotional Faces [CD-ROM]*. Stockholm: Karolinska Institutet, Department of Clinical Neuroscience, Psychology Section.

Mao, S. C., Hsiao, Y. H., & Gean, P. W. (2006). Extinction training in conjunction with a partial agonist of the glycine site on the nmda receptor erases memory trace. *Journal of Neuroscience, 26*(35), 8892-8899.

Maren, S. (1999). Long-term potentiation in the amygdala: A mechanism for emotional learning and memory.[see comment]. *Trends in Neurosciences, 22*(12), 561-567.

Maren, S. (2001). Neurobiology of pavlovian fear conditioning. *Annual Review of Neuroscience, 24*, 897-931.

Maren, S., & Chang, C. H. (2006). Recent fear is resistant to extinction. *Proceedings of the National Academy of Sciences of the United States of America, 103*(47), 18020-18025.

Maren, S., & Fanselow, M. S. (1995). Synaptic plasticity in the basolateral amygdala induced by hippocampal formation stimulation in vivo. *Journal of Neuroscience, 15*(11), 7548-7564.

Maren, S., & Quirk, G. J. (2004). Neuronal signalling of fear memory. *Nature Reviews Neuroscience, 5*(11), 844-852.

Mariano de Bitencourt, R., Pamplona, F. A., & Takahashi, R. N. (2013). A current overview of cannabinoids and glucocorticoids in facilitating extinction of aversive memories: Potential extinction enhancers. *Neuropharmacology, 64*, 389-395.

Marsicano, G., Wotjak, C. T., Azad, S. C., Bisogno, T., Rammes, G., Cascio, M. G., . . . Lutz, B. (2002). The endogenous cannabinoid system controls extinction of aversive memories. *Nature, 418*(6897), 530-534.

Martin, S. J., Grimwood, P. D., & Morris, R. G. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience, 23*, 649-711.

McDonald, A. J., Mascagni, F., & Guo, L. (1996). Projections of the medial and lateral prefrontal cortices to the amygdala: A phaseolus vulgaris leucoagglutinin study in the rat. *Neuroscience, 71*(1), 55-75.

McGaugh, J. L. (2000). Neuroscience - memory - a century of consolidation. *Science, 287*(5451), 248-251.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27*, 1-28.

McNally, G. P., & Westbrook, R. F. (2006). Predicting danger: The nature, consequences, and neural mechanisms of predictive fear learning. *Learning & Memory, 13*(3), 245-253.

McSweeney, F. K., & Swindell, S. (2002). Common processes may contribute to extinction and habituation. *Journal of General Psychology, 129*(4), 364-400.

Milad, M.R., Orr, S.P., Pitman, R.K., & Rauch, S.L. (2005). Context modulation of memory for fear extinction in humans. *Psychophysiology*, *42*(4), 456-464.

Milad, M. R., Quinn, B. T., Pitman, R. K., Orr, S. P., Fischl, B., & Rauch, S. L. (2005). Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. *Proceedings of the National Academy of Sciences of the United States of America, 102*(30), 10706-10711.

Milad, M. R., & Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature, 420*(6911), 70-74.

Milad, M. R., Vidal-Gonzalez, I., & Quirk, G. J. (2004). Electrical stimulation of medial prefrontal cortex reduces conditioned fear in a temporally specific manner. *Behavioral Neuroscience, 118*(2), 389-394.

Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological Psychiatry, 62*(5), 446-454.

Mineka, S., & Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of Experimental Psychology-General, 122*(1), 23-38.

Mineka, S., Davidson, M., Cook, M., & Keir, R. (1984). Observational conditioning of snake fear in rhesus-monkeys. *Journal of Abnormal Psychology, 93*(4), 355-372.

Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsie shock after reactivation of a consolidated memory trace. *Science, 160*(3827), 554-&.

Miserendino, M. J., Sananes, C. B., Melia, K. R., & Davis, M. (1990). Blocking of acquisition but not expression of conditioned fear-potentiated startle by nmda antagonists in the amygdala. *Nature, 345*(6277), 716-718.

Monfils, M. H., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: Key to persistent attenuation of fear memories. *Science, 324*(5929), 951-955.

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature, 393*(6684), 467-470.

Morris, R. W., & Bouton, M. E. (2007). The effect of yohimbine on the extinction of conditioned fear: A role for context. *Behavioral Neuroscience, 121*(3), 501-514.

Moustafa, A. A., Gilbertson, M. W., Orr, S. P., Herzallah, M. M., Servatius, R. J., & Myers, C. E. (2013). A model of amygdala–hippocampal–prefrontal interaction in fear conditioning and extinction in animals. *Brain and Cognition, 81*, 29-43.

Myers, K. M., & Davis, M. (2007). Mechanisms of fear extinction. *Molecular Psychiatry, 12*(2), 120-150.

Myers, K. M., Ressler, K. J., & Davis, M. (2006). Different mechanisms of fear extinction dependent on length of time since fear acquisition. *Learning & Memory, 13*(2), 216-223.

Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews Neuroscience, 10*(3), 224-234.

Nader, K., Schafe, G. E., & LeDoux, J. E. (2000a). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature, 406*(6797), 722-726.

Nader, K., Schafe, G. E., & LeDoux, J. E. (2000b). The labile nature of consolidation theory. *Nature Reviews Neuroscience, 1*(3), 216-219.

Norrholm, S. D., Jovanovic, T., Vervliet, B., Myers, K. M., Davis, M., Rothbaum, B. O., & Duncan, E. J. (2006). Conditioned fear extinction and reinstatement in a human fear-potentiated startle paradigm. *Learning & Memory, 13*(6), 681-685.

Norrholm, S. D., Vervliet, B., Jovanovic, T., Boshoven, W., Myers, K. M., Davis, M., . . . Duncan, E. J. (2008). Timing of extinction relative to acquisition: A parametric analysis of fear extinction in humans. *Behavioral Neuroscience, 122*(5), 1016-1030.

Öhman, A. (1979). The orienting response, attention, and learning: An information-processing perspective. In E. H. v. O. H. D. Kimmel, & J. F. Orlebeke (Ed.), *The orienting reflex in humans* (pp. 443-471). Hillsdale, N. J: Erlbaum

Öhman, A. (2000). Encyclopedia of stress. In G. Fink (Ed.), (Vol. 2, pp. 111-116). San Diego: Academic Press.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483-522.

Öhman, A., & Soares, J. J. F. (1993). On the automatic nature of phobic fear - conditioned electrodermal responses to masked fear-relevant stimuli. *Journal of Abnormal Psychology, 102*(1), 121-132.

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: The neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience, 2*(1), 3-11.

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after pavlovian, observational, and instructed fear. *Psychological Science, 15*(12), 822-828.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience, 10*(9), 1095-1102.

Oyarzun, J. P., Lopez-Barroso, D., Fuentemilla, L., Cucurell, D., Pedraza, C., Rodriguez-Fornells, A., & de Diego-Balaguer, R. (2012). Updating fearful memories with extinction training during reconsolidation: A human study using auditory aversive stimuli. *PLoS ONE, 7*(6).

Pare, D., Quirk, G. J., & LeDoux, J. E. (2004). New vistas on amygdala networks in conditioned fear. *Journal of Neurophysiology, 92*(1), 1-9.

Parnas, A. S., Weber, M., & Richardson, R. (2005). Effects of multiple exposures to d-cycloserine on extinction of conditioned fear in rats. *Neurobiology of Learning & Memory, 83*(3), 224-231.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex.* . London: Oxford University Press.

Pearce, J. M., & Hall, G. (1980). A model for pavlovian learning - variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*(6), 532-552.

Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction learning in humans: Role of the amygdala and vmpfc. *Neuron, 43*(6), 897-905.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron, 48*(2), 175-187.

Pitkanen, A., Savander, V., & LeDoux, J. E. (1997). Organization of intra-amygdaloid circuitries in the rat: An emerging framework for understanding functions of the

amygdala.[see comment][erratum appears in trends neurosci 1998 feb;21(2):52]. *Trends in Neurosciences, 20*(11), 517-523.

Plendl, W., & Wotjak, C. T. (2010). Dissociation of within- and between-session extinction of conditioned fear. *Journal of Neuroscience, 30*(14), 4990-4998.

Ponnusamy, R., Nissim, H. A., & Barad, M. (2005). Systemic blockade of d2-like dopamine receptors facilitates extinction of conditioned fear in mice. *Learning & Memory, 12*(4), 399-406.

Prenoveau, J. M., Craske, M. G., Liao, B., & Ornitz, E. M. (2012). Human fear conditioning and extinction: Timing is everything…or is it? *Biological Psychology, doi:10.1016/j.biopsycho.2012.02.005*.

Quirk, G. J., Likhtik, E., Pelletier, J. G., & Pare, D. (2003). Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. *Journal of Neuroscience, 23*(25), 8800-8807.

Quirk, G. J., Repa, C., & LeDoux, J. E. (1995). Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: Parallel recordings in the freely behaving rat. *Neuron, 15*(5), 1029-1039.

Quirk, G.J., Russo, G.K., Barron, J.L., & Lebron, K. (2000). The role of the ventromedial prefrontal cortex in the recovery of extinguished fear. *Journal of Neuroscience*, *20*(16), 6225-31.

Rachman, S. (1989). The return of fear - review and prospect. *Clinical Psychology Review, 9*(2), 147-168.

Rauch, S. L., Shin, L. M., & Phelps, E. A. (2006). Neurocircuitry models of posttraumatic stress disorder and extinction: Human neuroimaging research--past, present, and future. *Biological Psychiatry, 60*(4), 376-382.

Repa, J. C., Muller, J., Apergis, J., Desrochers, T. M., Zhou, Y., & LeDoux, J. E. (2001). Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nature Neuroscience, 4*(7), 724-731.

Rescorla, R.A., & Wagner. A.R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. B. W. F. Prokasy (Ed.), *Classical conditioning: Ii. Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review, 74*(1), 71-&.

Rescorla, R. A. (1988). Pavlovian conditioning - its not what you think it is. *American Psychologist, 43*(3), 151-160.

Rescorla, R. A. (2004). Spontaneous recovery. *Learning & Memory, 11*(5), 501-509.

Rescorla, R. A., & Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus *Journal of Experimental Psychology, 104*(1), 88-96.

Ressler, K. J., Rothbaum, B. O., Tannenbaum, L., Anderson, P., Graap, K., Zimand, E., . . . Davis, M. (2004). Cognitive enhancers as adjuncts to psychotherapy: Use of d-cycloserine in phobic individuals to facilitate extinction of fear. *Archives of General Psychiatry, 61*(11), 1136-1144.

Ritter, B. (1968). Group desensitizationof childrens snake phobias using vicarious and contact desensitization procedures. *Behaviour Research and Therapy, 6*(1).

Robbins, S. J. (1990). Mechanisms underlying spontaneous-recovery in autoshaping. *Journal of Experimental Psychology-Animal Behavior Processes, 16*(3), 235-249.

Rogan, M. T., Staubli, U. V., & LeDoux, J. E. (1997). Fear conditioning induces associative long-term potentiation in the amygdala. *Nature, 390*(6660), 604-607.

Rosenkranz, J. A., Moore, H., Grace, A. A. (2003). The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. *Journal of Neuroscience, 23*(35), 11054-11064.

Santini, E., Ge, H., Ren, K., Pena de Ortiz, S., & Quirk, G. J. (2004). Consolidation of fear extinction requires protein synthesis in the medial prefrontal cortex. *Journal of Neuroscience, 24*(25), 5704-5710.

Santini, E., Muller, R. U., & Quirk, G. J. (2001). Consolidation of extinction learning involves transfer from nmda-independent to nmda-dependent memory. *Journal of Neuroscience, 21*(22), 9009-9017.

Sara, S. J. (2000). Retrieval and reconsolidation: Toward a neurobiology of remembering. *Learning & Memory, 7*(2), 73-84.

Schafe, G. E., & LeDoux, J. E. (2000). Memory consolidation of auditory pavlovian fear conditioning requires protein synthesis and protein kinase a in the amygdala. *Journal of Neuroscience, 20*(18), art. no.-RC96.

Schiller, D., Cain, C. K., Curley, N. G., Schwartz, J. S., Stern, S. A., LeDoux, J. E., & Phelps, E. A. (2008). Evidence for recovery of fear following immediate extinction in rats and humans. *Learning & Memory, 15*(6), 394-402.

Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature, 463*(7277), 49-U51.

Schmajuk, N. A., & Moore, J. W. (1989). Effects of hippocampal manipulations on the classically-conditioned nictitating-membrane response - simulations by an attentional-associative model. *Behavioural Brain Research, 32*(2), 173-189.

Sehlmeyer, C., Schoning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., & Konrad, C. (2009). Human fear conditioning and extinction in neuroimaging: A systematic review. *PLoS ONE, 4*(6).

Seligman, L. D., & Wuyek, L. A. (2005). Vicarious extinction In M. Hersen (Ed.), *Encyclopedia of behavior modification and cognitive behavior therapy, vol iii: Educational applications (v. 1)*

Shipley, R. H. (1974). Extinction of conditioned fear in rats as a function of several parameters of cs exposure. *Journal of Comparative and Physiological Psychology, 87*, 699–707.

Sierra-Mercado, D. Jr., Corcoran, K. A., Lebron-Milad, K., Quirk, G. J. (2006). Inactivation of the ventromedial prefrontal cortex reduces expression of conditioned fear and impairs subsequent recall of extinction. *European Journal of Neuroscience, 24*(6), 1751-1758.

Smith, Y., Pare, J. F., & Pare, D. (2000). Differential innervation of parvalbumin-immunoreactive interneurons of the basolateral amygdaloid complex by cortical and intrinsic inputs. *Journal of Comparative Neurology, 416*(4), 496-508.

Soeter, M., & Kindt, M. (2010). Dissociating response systems: Erasing fear from memory. *Neurobiology of Learning and Memory, 94*(1), 30-41.

Soeter, M., & Kindt, M. (2011). Disrupting reconsolidation: Pharmacological and behavioral manipulations. *Learning & Memory, 18*(6), 357-366.

Soeter, M., & Kindt, M. (2012a). Erasing fear for an imagined threat event. *Psychoneuroendocrinology, 37*(11), 1769-1779.

Soeter, M., & Kindt, M. (2012b). Stimulation of the noradrenergic system during memory formation impairs extinction learning but not the disruption of reconsolidation. *Neuropsychopharmacology, 37*, 1204–1215.

Sotres-Bayon, F., Bush, D. E., & LeDoux, J.E. (2004). Emotional perseveration: An update on prefrontal-amygdala interactions in fear extinction. *Learning & Memory, 11*(5), 525-535.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (socr): A formalization of the comparator hypothesis. *Psychological Review, 114*(3), 759-783.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks - expectation and prediction. *Psychological Review, 88*(2), 135-170.

Suzuki, A., Josselyn, S. A., Frankland, P. W., Masushige, S., Silva, A. J., & Kida, S. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *Journal of Neuroscience, 24*(20), 4787-4795.

Thompson, R. F., & Spencer, W. A. (1966). Habituation - a model phenomenon for study of neuronal substrates of behavior. *Psychological Review, 73*(1), 16-&.

Walker, D. L., & Davis, M. (2002). The role of amygdala glutamate receptors in fear learning, fear-potentiated startle, and extinction. *Pharmacology, Biochemistry & Behavior, 71*(3), 379-392.

Walker, D. L., Ressler, K. J., Lu, K. T., & Davis, M. (2002). Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of d-cycloserine as assessed with fear-potentiated startle in rats. *Journal of Neuroscience, 22*(6), 2343-2351.

van Well, S., Visser, R., Scholte, H., & Kindt, M. (2012). Neural substrates of individual differences in human fear learning: Evidence from concurrent fmri, fear-potentiated startle, and us-expectancy data. *Cognitive, Affective, & Behavioral Neuroscience, 12*(3), 499-512.

Wang, S. H., Alvares, L. D., & Nader, K. (2009). Cellular and systems mechanisms of memory strength as a constraint on auditory fear reconsolidation. *Nature Neuroscience, 12*(7), 905-U116.

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquistion context. *Behaviour Research and Therapy, 43*(3), 323-336.

Weike, A. I., Hamm, A. O., Schupp, H. T., Runge, U., Schroeder, H. W. S., & Kessler, C. (2005). Fear conditioning following unilateral temporal lobectomy: Dissociation of conditioned startle potentiation and autonomic learning. *Journal of Neuroscience, 25*(48), 11117-11124.

Vertes, R. P. (2004). Differential projections of the infralimbic and prelimbic cortex in the rat. *Synapse, 51*(1), 32-58.

Wiens, S., & Öhman, A. (2007). Probing unconscious emotional processes: On becoming a successful masketeer. In J. A. Coan & J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 65-91). New York: Oxford University Press.

Wilensky, A. E., Schafe, G. E., Kristensen, M. P., & LeDoux, J. E. (2006). Rethinking the fear circuit: The central nucleus of the amygdala is required for the acquisition, consolidation, and expression of pavlovian fear conditioning. *Journal of Neuroscience, 26*(48), 12387-12396.

Woods, A. M., & Bouton, M. E. (2006). D-cycloserine facilitates extinction but does not eliminate renewal of the conditioned emotional response. *Behavioral Neuroscience, 120*(5), 1159-1162.

Woods, A. M., & Bouton, M. E. (2008). Immediate extinction causes a less durable loss of performance than delayed extinction following either fear or appetitive conditioning. *Learning & Memory, 15*(12), 909-920.

Yang, Y. L., Chao, P. K., & Lu, K. T. (2006). Systemic and intra-amygdala administration of glucocorticoid agonist and antagonist modulate extinction of conditioned fear. *Neuropsychopharmacology, 31*(5), 912-924.

Yeo, A. G. (1974). Acquisition of conditioned suppression as a function of interstimulus interval duration. *Quarterly Journal of Experimental Psychology, 26*, 405-416.

Zushida, K., Sakurai, M., Wada, K., & Sekiguchi, M. (2007). Facilitation of extinction learning for contextual fear memory by pepa: A potentiator of ampa receptors. *Journal of Neuroscience, 27*(1), 158-166.