A translational perspective on neural circuits of fear extinction: Current promises and challenges

Dieuwke Sevenster[a,b,*], Renée M. Visser[c], Rudi D'Hooge[a]

[a] Laboratory of Biological Psychology, Department of Psychology, KU Leuven, Tiensestraat 102, B-3000 Leuven, Belgium
[b] Clinical Psychology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands
[c] Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

ABSTRACT

Fear extinction is the well-known process of fear reduction through repeated re-exposure to a feared stimulus without the aversive outcome. The last two decades have witnessed a surge of interest in extinction learning. First, extinction learning is observed across species, and especially research on rodents has made great strides in characterising the physical substrate underlying extinction learning. Second, extinction learning is considered of great clinical significance since it constitutes a crucial component of exposure treatment. While effective in reducing fear responding in the short term, extinction learning can lose its grip, resulting in a return of fear (i.e., laboratory model for relapse of anxiety symptoms in patients). Optimization of extinction learning is, therefore, the subject of intense investigation. It is thought that the success of extinction learning is, at least partly, determined by the mismatch between what is expected and what actually happens (prediction error). However, while much of our knowledge about the neural circuitry of extinction learning and factors that contribute to successful extinction learning comes from animal models, translating these findings to humans has been challenging for a number of reasons. Here, we present an overview of what is known about the animal circuitry underlying extinction of fear, and the role of prediction error. In addition, we conducted a systematic literature search to evaluate the degree to which state-of-the-art neuroimaging methods have contributed to translating these findings to humans. Results show substantial overlap between networks in animals and humans at a macroscale, but current imaging techniques preclude comparisons at a smaller scale, especially in sub-cortical areas that are functionally heterogeneous. Moreover, human neuroimaging shows the involvement of numerous areas that are not typically studied in animals. Results obtained in research aimed to map the extinction circuit are largely dependent on the methods employed, not only across species, but also across human neuroimaging studies. Directions for future research are discussed.

1. Introduction

Aversive experiences often leave a mark: seeing a dog shortly after being bitten by one may trigger a range of defensive responses, such as heavy perspiration, trembling, freezing, and the tendency to avoid the animal. Yet, with repeated encounters with harmless dogs such responses usually fade, a process referred to as 'fear extinction'. The process of extinction has been studied for almost a century, starting with the classic studies by Ivan Pavlov (Pavlov, 1927). Today, fear extinction is a hot topic in neuroscience given the tremendous progress in uncovering its neural basis and its clinical significance.

Extinction is preceded by fear conditioning, the process of pairing a neutral stimulus with an aversive outcome (or threat conditioning as some prefer to call it in animals LeDoux, 2014). The formation of these

associations during fear acquisition is very rapid and extremely robust. This is in stark contrast with fear extinction learning, the decrement of responding through unreinforced presentation of the conditioned stimulus (CS), which requires extensive training and results in a relatively fragile memory trace (Bouton, 1993, 2002, 2004). Extinction training produces a new memory trace that inhibits the original association of the CS with the unconditioned stimulus (US). Maintaining the inhibitory memory is thus critical for long-term retention of extinction. This remains a great challenge as fear returns under a variety of conditions such as re-exposure to the US (reinstatement), the passage of time (spontaneous recovery), and post-extinction context change (i.e., renewal) (Bouton, 2002). Here, we use the term 'extinction' exclusively to refer to extinction of conditioned fear.

By and large, most of our knowledge about the principles of fear

* Corresponding author at: Utrecht University, Department of Clinical Psychology, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands.
  E-mail address: d.sevenster@uu.nl (D. Sevenster).

extinction is derived from a vast body of animal literature. Notably, many affective responses appear to be highly conserved between different non-human mammalian species, and many of the underlying cellular and molecular processes of neuroplasticity can even be examined in invertebrates such as insects and gastropods. Particularly the last two decades of neuroscientific research in animals have produced exciting new insights into the representation and the dynamic nature of emotional memory in general (Josselyn, Köhler, & Frankland, 2015; Tonegawa, Liu, Ramirez, & Redondo, 2015), and extinction specifically (e.g., Herry et al., 2008; Moscarello & Maren, 2018; Senn et al., 2014; Xu et al., 2016).

What makes insight into the neural basis of extinction so important is its clinical significance. Extinction learning is considered the crux of exposure therapy, during which patients suffering from anxiety or stress-disorders are confronted with the object of their fear in a safe environment. While exposure is effective in reducing anxiety symptoms, some patients respond insufficiently or experience a relapse of symptoms on the long run (Craske, 1999; Resick, Williams, Suvak, Monson, & Gradus, 2012). The return of fear responses provides a laboratory model of relapse of fear following treatment that is often observed in patients after exposure therapy (Eddy, Dutra, Bradley, & Westen, 2004; Resick et al., 2012). It is important to note that detection of danger and immediate engagement of the defensive system in the face of threat is crucial for survival. Learning to associate stimuli with danger can help an individual to optimize effective threat detection. Thus, both fear learning and appropriate return of fear responding following extinction are highly adaptive. Crucially, in patients the ability for adaptive fear learning and inhibition of inappropriate fear responding is impaired (Duits et al., 2015; Duits, Cath, Heitland, & Baas, 2016; Peri, Ben-Shakhar, Orr, & Shalev, 2000; VanElzakker, Kathryn Dahlgren, Caroline Davis, Dubois, & Shin, 2014).

Insights from animal studies on the extinction circuitry are indispensable in providing hypotheses about extinction learning and return of fear in humans. These hypotheses can be tested in laboratory settings, to help identify factors that increase the risk of relapse after successful exposure therapy (Eddy et al., 2004; Resick et al., 2012), as well as to guide the innovation and optimization of clinical treatments. However, the translation of our vast body of knowledge about animal extinction learning to the human condition remains a formidable task. Neuroimaging in humans has been suggested to provide a crucial translational step in assessing to what extent the neural circuitries underlying emotional memory systems are similar across species (Milad & Quirk, 2012; Sehlmeyer et al., 2009; Visser, Lau-Zhu, Henson, & Holmes, 2018). However, as will be discussed in Sections 3 and 4, drawing parallels between the animal and human extinction circuits using modern neuroimaging techniques is not always straightforward.

Here, we present a succinct overview of what is established knowledge about the animal circuitry underlying extinction of fear (Section 2), and compare this with recent findings from human imaging studies as identified by a systematic literature search (Section 3). In the discussion of both animal and human extinction circuitry we will consider the role of prediction error, that is, a mismatch between what is expected (e.g., CS-US) and what actually happens (e.g., CS-no US). In doing so, this review aims to provide both an update and extension of previous reviews (Milad & Quirk, 2012; Sehlmeyer et al., 2009). We will argue that translation of the animal extinction circuitry to the human brain is more complex than previously thought. We will discuss to what degree recent advances in human imaging and analyses techniques might aid to overcome some of these difficulties (Section 4), as well as highlight other avenues for future research.

2. Extinction network in rodents

Extinction is ethologically relevant and evolutionary adaptive, since it is part of an animal's behavioural repertoire that enables it to adjust its behaviour to a perpetually changing environment. The extinction of

fear occurs even in primitive organisms that possess only the most basic learning abilities. Also, more advanced invertebrates such as snails and insects display extinction when a previously encountered threat is no longer present (Eisenhardt, 2014). Studies in rodents such as mice and rats have been particularly important to our present understanding of the behavioural and neural mechanisms of extinction, because of the homologies between rodent and primate brain structure and function (Barad, 2005).

Generally, Pavlovian fear (or threat) conditioning research in rodents entails the coupling of a discrete cue (e.g., light or tone) or a context (CS), with foot shocks (US). Freezing response is the most frequently used readout of conditioned responding (CR) to the CS. Fear extinction is established by repeated presentation of the CS alone. Evidence of extinction learning can be assessed at the end of the extinction training (reduction from beginning to end) or at a later time point, usually 24 h later. Assessing extinction learning at a later time point is referred to as *extinction recall*. Within-session extinction is not considered a good marker for a longer-term reduction in fear responding, or in therapeutic terms, of 'successful' extinction learning (Craske, 2015; Plendl & Wotjak, 2010): individuals that show a strong reduction of fear responses from the beginning to the end of a session may show a substantial or complete recovery of conditioned fear at the beginning of the next session, while individuals with a more moderate reduction of fear may show better retention of this reduction over time. Extinction *recall*, a reduction of fear that persists, may thus be a better, and clinically more relevant, indicator of successful extinction. The standard model of Pavlovian conditioning is highly relevant for normal or adaptive fear acquisition and extinction. Extensions of the model that increase the unpleasantness of the learning situation, for example using higher shock intensity or a prior stress episode, could potentially more accurately model pathological fear (Desmedt, Marighetto, & Piazza, 2015; Izquierdo, Wellman, & Holmes, 2006). Furthermore, we note a recent and increasing interest in the empirical characterization of heterogeneity in fear learning, extinction and return of fear in animal models (e.g. Galatzer-Levy, Bonanno, Bush, & Ledoux, 2013; Shumake, Jones, Auchter, & Monfils, 2018), which is expected to push the field forward in terms of identifying factors contributing to mental health disorders and markers to predict treatment response. For pragmatic reasons, we restrict the current perspective to research using 'standard' models of fear, and by focusing more on average responses than on heterogeneity.

Several procedures are used to uncover the return of fear following extinction training: a return of freezing response to the CS after the passage of time (e.g. one week) indicates spontaneous recovery; reinstatement of fear to the CS is observed after unsignaled presentation of the foot shock; a renewal of fear occurs upon a switch from the context in which extinction (context B) took place to the original acquisition context (context A) or a new context (context C). To this end, environmental features (shape; odour; lighting) of the cage in which fear conditioning, extinction, and renewal test take place are manipulated. Renewal experiments clearly demonstrate the contextual dependency of extinction (see Section 2.2) (Bouton, 2002, 2004).

2.1. Prelimbic-amygdala and Infralimbic-amygdala projections

Decades of research have shown that memory for extinction is not situated in a single brain area. Rather, it is distributed across a network which main structures include the amygdala, hippocampus and prefrontal cortex (PFC). This network has been studied extensively in animals and is addressed only concisely here (for review see e.g., Moscarello & Maren, 2018; Orsini & Maren, 2012; Quirk & Mueller, 2008). The key circuitry is depicted schematically in Fig. 1. In short, the amygdala has been proposed to play a crucial role in the acquisition and expression of conditioned fear responding. The central nucleus of the amygdala (CEA) controls the expression of conditioned fear. Projections from the basolateral complex of the amygdala (BLA) modulate CEA
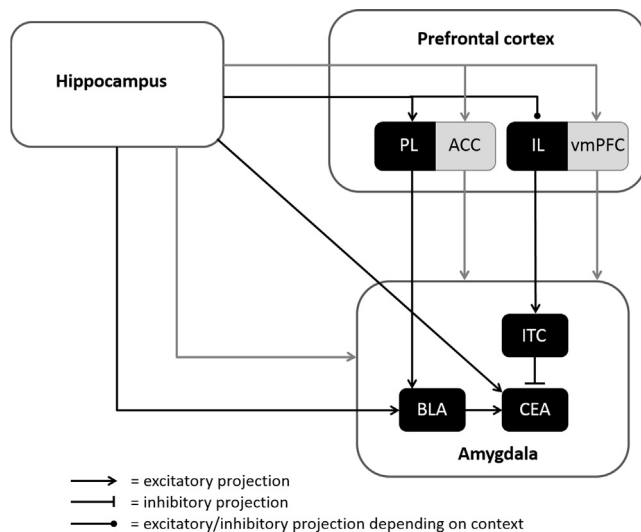
Fig. 1. Neural networks of extinction. Neural networks for extinction have been investigated heavily in rodents and a network including at least the amygdala, hippocampus, and prefrontal areas has been identified. Substantial progress has been made in uncovering the subregions involved in extinction in rodents; the prelimbic area (PL) innervates the basolateral amygdala (BLA), which in turn projects to the central amygdala (CEA). The CEA controls conditioned responding and receives input from the infralimbic area (IL), mediated through the intercalated cells (ITC) of the amygdala. The hippocampus projects directly, and indirectly via the PL and IL, to the amygdala. In humans, it has been hypothesized that the same areas are involved in extinction learning. The anterior cingulate cortex (ACC) and the ventromedial prefrontal cortex (vmPFC) are assumed to constitute the human homologue of the PL and IL, respectively. Although the amygdala is generally assumed to be involved in fear learning and extinction in humans, neuroimaging evidence for this involvement is scarce. White areas constitute both the animal and human extinction network; black areas are specifically identified in animals; grey areas are related to the human extinction network.

activity (LeDoux, 2000; Vidal-Gonzalez, Vidal-Gonzalez, Rauch, & Quirk, 2006). The PFC exerts a dual control in expression of conditioned fear, as it can both facilitate and inhibit fear responding, which depends on different amygdala connections (Senn et al., 2014). Prelimbic (PL) PFC sends excitatory projections to the basal amygdala (BA), which in turn innervates the central amygdala (CEA) resulting in expression of conditioned fear (LeDoux, 2000; Vidal-Gonzalez et al., 2006) (Fig. 1).

Extinction learning is thought to heavily depend on the infralimbic area (IL) of the PFC. Lesion studies showed that extinction depends on IL neuroplasticity and the concomitant formation of novel or updated memory networks (Lebron, Milad, & Quirk, 2004; Quirk, Russo, Barron, & Lebron, 2000). Manipulating the processes of neural plasticity, especially those involved in memory consolidation and updating, affected extinction. For example, a series of experiments showed that infusion of brain-derived neurotrophic factor (BDNF) – a protein that controls neural growth and survival (Bramham & Messaoudi, 2005) – into the IL cortex of fear-conditioned rats produced an extinction-like effect without affecting the original fear memory (Peters, Dieppa-Perea, Melendez, & Quirk, 2010). More specifically, BDNF infusion 60 min before extinction training significantly reduced fear responding during extinction training and on subsequent tests of extinction recall, reversed naturally occurring extinction deficits, and even reduced fear responding in the absence of further extinction training. This effect was specific to inhibition of conditioned fear, as open-field anxiety was still intact and BDNF infusion prior to conditioning did not reduce subsequent fear responding. Yet, this effect was transient as conditioned fear responses could be reinstated, indicating that BDNF infusion did not degrade the original fear memory. Further experiments

demonstrated that the hippocampus (rather than amygdala or IL) is the source of IL BDNF, and that similar extinction-like effects could be achieved by boosting hippocampal BDNF, unless a BDNF-inactivating antibody was co-administered in the IL, indicating that the IL mPFC is the primary site of action for hippocampal BDNF (Peters et al., 2010). These experiments showed that the IL is crucial for inhibitory extinction learning. Other studies showed that IL exerts this inhibition through projections to the amygdala. This inhibition is not due to direct IL-CEA connections, but is mediated by the intercalated cells (ITC) of the amygdala that inhibit the CEA (Amano, Unal, & Paré, 2010). In addition, it is thought that the IL has reciprocal connections to the BLA. Although the BLA is mostly recognized for its role in the acquisition of fear responses, this is most likely not its only role. An optogenetic study showed a dissociation between fear neurons and extinction neurons within the BA (Herry et al., 2008). These extinction neurons were responsive only to extinguished stimuli and were connected to the mPFC, possibly affecting activity in the CE via ITC or direct projections (Orsini & Maren, 2012). Understanding of direct and indirect IL-amygdala connections is developing rapidly and cannot be fully addressed here. What is important is that the net effect of these IL-amygdala projections is reduced responsiveness of CEA to BLA input, resulting in reduced fear responding (Quirk, Likhtik, Pelletier, & Paré, 2003).

2.2. Hippocampus

The hippocampus exerts a dual role in extinction learning and recall; depending on the conditions the hippocampus activates or inhibits fear expression. Hippocampal involvement during extinction learning has been considered a prerequisite for extinction recall; inactivation of the hippocampus during extinction training resulted in impaired extinction recall; fear was expressed both in the context in which extinction took place and other contexts (Corcoran, Desmond, Frey, & Maren, 2005; Sierra-Mercado, Padilla-Coreano, & Quirk, 2011).

Extinction learning and recall is highly context dependent (Bouton, 2002, 2004); extinction is best recalled in the context in which it took place and a switch to another context uncovers the conserved relation of the CS and its aversive consequence. The hippocampus is thought critical for establishing a context representation during extinction learning, and the regulation of fear renewal following a context switch (Maren, Phan, & Liberzon, 2013). Down-regulation of the hippocampus during extinction recall facilitates generalization of extinction to other contexts. Promisingly, pharmacological inactivation of the hippocampus (Corcoran & Maren, 2001) or specific optogenetic inhibition during extinction recall prevented fear renewal (Xu et al., 2016). The role of the hippocampus in activation or inhibition of fear expression appears to be controlled by direct and indirect pathways. There are direct projections from the hippocampus to the BLA, and indirect projections through the IL and PL (Orsini, Kim, Knapska, & Maren, 2011; Orsini & Maren, 2012) (Fig. 1). Pathways to the BLA and PL are thought to be involved in fear expression; fear renewal was indeed shown to recruit hippocampal-BLA projections (Jin & Maren, 2015; Orsini et al., 2011). Recently, in addition to projection to the BLA, weaker hippocampal projections to the CEA have been identified (Xu et al., 2016). Different populations of CA1 hippocampal neurons target the BLA and CEA and specific optogenetic inhibition of the hippocampal-CEA pathway during renewal test prevented fear renewal (Xu et al., 2016). Moreover, exciting recent insights demonstrated that hippocampus-mediated inhibition of IL could be central to fear relapse; activation of hippocampal input into amygdala-projecting IL neurons resulted in a relapse of fear (in the extinction context), while inactivation of hippocampal-IL pathway prevented renewal (Marek et al., 2018).

In sum, the hippocampus mediates both fear activation and inhibition. Hippocampal activity during extinction is associated with extinction plasticity. Hippocampal inactivation during extinction recall may facilitate extinction generalization to other contexts.

2.3. Prediction error in rodents

Fear extinction entails adaptation to changing environmental demands; when the CS is no longer a good predictor of threat, it should no longer be considered as unsafe. An important factor in the effective adjustment to the reversal in contingencies is mediated by a violation of expectancies. The greater the mismatch between what is expected (CS-US) and what actually happens (CS-no US), the greater the opportunity for new learning (Rescorla & Wagner, 1972). While prediction error is considered the driving force of fear extinction, it receives little attention in the traditional conceptualization of the neural circuitry of extinction.

Following fear acquisition, unreinforced CS exposure can have opposite behavioural outcomes: initially, the CS leads to mere retrieval of the original fear memory, leading to anticipation of threat and expression of conditioned fear. With repeated presentation of the CS without the US, it may no longer be considered as unsafe, and conditioned fear responding may be inhibited. How does the brain determine when to continue to anticipate threat, or when to adapt to changing environmental demands and transition to safety behaviour? Prediction of the US and confirmation or violation of prediction might control which network (expression or inhibition of fear) is engaged. Rescorla and Wagner (1972) argued that the greater the mismatch between what is expected (based on previous experience e.g., CS-US) and what actually occurs (CS-no US), the greater the opportunity for new learning (but alternative models have been used as well; Delamater & Westbrook, 2014). While unexpected reinforcement (positive PE) drives fear learning, absence of expected reinforcement (negative PE) is linked to extinction learning, and an extinction memory trace is formed when the actual outcome on a learning trial is less than expected. Note that this model (Rescorla & Wagner, 1972) traditionally regarded extinction as unlearning. The notion that extinction consists of additional inhibitory learning is later accounted for by the observation of return of fear following extinction (Bouton, 2002, 2004; see also introduction). While prediction error is considered the driving force of fear extinction, it receives little attention in the traditional conceptualization of the neural circuitry of extinction. Animal networks of PE will be discussed below.

2.3.1. Prediction error and the hippocampus

The hippocampus is explicitly associated with contextual processing during and following extinction (see Sections 2 and 3.3). It is, however, likely that the role of the hippocampus in extinction extends beyond contextual modulation. Research from a different field recognizes the hippocampus as an essential region in error processing (i.e., matching current incoming information with predictions). Novelty detection and prediction error signaling is associated with the CA1 region of the hippocampus (Barbeau, Chauvel, Moulin, Regis, & Liégeois-Chauvel, 2017; Giovannini et al., 2001; Huh et al., 2009; Lisman & Grace, 2005; Lisman & Otmakhova, 2001; Radulovic & Tronson, 2010; Tronson et al., 2009). Note that novelty and prediction error are related but distinct concepts. For example, a novel stimulus that is not followed by a threatening event may generate less prediction error than a familiar threatening stimulus that is not reinforced. Thus, it may not be the overall novelty but rather the amount of PE drives learning. Alternatively, PE alone may be insufficient to explain some of the reported results regarding extinction, and additional novelty-driven attentional mechanisms may need to be taken into account (e.g. Larrauri & Schmajuk, 2008).

Different molecular markers of novelty and PE have been identified; generally, expression of cFos in CA1 increases in response to novelty (Matsuo & Mayford, 2008; Monfil et al., 2018; VanElzakker, Fevurly, Breindel, & Spencer, 2008). Specific for fear extinction is PE-induced activation of phospho extracellular signal-regulated kinases (pErk) in CA1 (Huh et al., 2009; Radulovic & Tronson, 2010; Tronson et al., 2009). Given the theoretical emphasis on PE in extinction it is surprising that, to our knowledge, little more research has been conducted

in this area.

Altogether, the hippocampus likely plays a dual role in extinction, possibly being crucial for both contextual modulation and error processing. The CA1 may constitute the hippocampal locus of these processes (Huh et al., 2009; Xu et al., 2016). Future studies could differentiate CA1 cell populations that are selectively involved in one of two processes. For example, optogenetic inhibition of those cells involved in contextual modulation should prevent fear renewal, while inhibition of PE-related cells should prevent extinction recall.

2.3.2. Prediction error and the ventral tegmental area (VTA)

In models of appetitive conditioning, PE has been frequently linked to dopaminergic signaling. An influential theory poses that there is a functional loop between the hippocampus and the ventral tegmental area (VTA) that controls memory encoding (Lisman & Grace, 2005); hippocampal novelty signals project to the VTA, and dopaminergic signals project back from the VTA to the hippocampus. Dopaminergic innervation of the hippocampus then leads to enhanced memory plasticity. Indeed, violation of prediction leads to alterations in firing rate of these dopaminergic systems; positive PE results in an increase, whereas negative PE results in a depression of dopamine firing (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007; Schultz, 2010; Waelti, Dickinson, & Schultz, 2001). It is important to note that while this pattern of activity is observed during appetitive conditioning, the decrease in dopamine associated with negative PE has rarely been observed in aversive conditioning (Li & McNally, 2014). Indeed, some authors have argued that the absence of expected negative reinforcement during extinction can be regarded as an appetitive prediction error (Raczka et al., 2011). Also, future rodent studies will have to characterize the role of the hippocampal-VTA loop in extinction. Nevertheless, the role of dopamine in fear learning and extinction has received considerable attention (for review see Abraham, Neve, & Lattal, 2014). D1 receptor knock-out mice showed delayed extinction learning (El-Ghundi, O'Dowd, & George, 2001). Impaired extinction learning was also observed when dopamine loss was restricted to the mPFC (Fernandez Espejo, 2003; Morrow, Elsworth, Rasmusson, & Roth, 1999), and a D1/D5 antagonist infusion in the IL resulted in impaired extinction retention (Hikind & Maroun, 2008). In contrast, selective activation of dopamine cells in the substantia nigra and subsequent D1 signaling in the striatum prevented fear renewal (Bouchet et al., 2018). Understanding the role of dopamine in associative learning is complicated given the existence of various dopamine receptor subtypes that in turn affect different signaling cascades. Apart from D1-like receptors (comprising D1/D5 receptors), the D2 receptor family (including D2/D3/D4 receptors) has also been implicated in extinction. Administration of D2 agents during extinction learning yielded conflicting results. For example, blocking D2 activity systemically improved extinction (Ponnusamy, Nissim, & Barad, 2005), whereas a D2 antagonist, administered systemically, to the nucleus accumbens (Holtzman-Assif, Laurent, & Westbrook, 2010), or to the IL prevented extinction (Mueller, Bravo-Rivera, & Quirk, 2010). Hence, while the exact role of dopamine in PE-driven extinction learning remains to be elucidated, studies thus far suggest that the circuitry involved extends beyond the hippocampal-VTA loop.

In sum, the extinction circuitry in rodents consists of PL-amygdala pathways that control the expression of learned fear, IL-amygdala pathways that mediate inhibition of conditioned responding, and hippocampal projections that control the contextual modulation of extinction. Circuitries involved in PE processing may include, but are not restricted to, the CA1 area of hippocampus and VTA dopaminergic projections.

3. The human extinction circuit

The typical fear conditioning and extinction paradigm in humans consists of pairing a picture (CS) with an aversive consequence (US).

The US usually entails electrical stimulation to the wrist, individually adjusted to a level that is uncomfortable but not painful. Less often, an auditory US is used, such as a loud scream. There are obvious ethical constraints on ecologically validity of fear conditioning and extinction in humans. Yet, within the limits of what is ethically possible, a number of variations of the paradigm have been shown to increase ecological validity. For example, use of fear relevant (e.g., spider) instead of neutral CS pictures (e.g., square, circle) results in more robust fear conditioning that is more resistant to extinction (Dawson, Schell, & Banis, 1986). More importantly, the use of virtual reality provides a great leap forward in terms of ecological validity. In these paradigms participants navigate through different virtual environments (e.g., room, park, street) in which discrete cues (e.g., tone or spider) are followed by electrical stimulation (Alvarez, Johnson, & Grillon, 2007; Huff et al., 2011). An advantage of the study of humans is that in addition to psychophysiological measurements (fear potentiated startle; skin conductance response, pupil dilation responses), subjective indices of conditioned responding (US-expectancy, distress) can be collected. As will be discussed, this may be particularly valuable in the context of assessing the role of prediction error (Section 3.5.1).

Compared to animals, it is more challenging to study the neural mechanisms underlying extinction in humans. While imaging studies can reveal regional (de)activation during extinction, inter-regional connectivity is difficult to establish, especially in deep brain structures such as the amygdala. Nevertheless, some similarities have been observed between the rodent and human extinction circuitry. The basic model proposes that, homologous to rodents, the human extinction network involves the amygdala, prefrontal cortex, and hippocampus. Most knowledge on this circuitry in humans is derived from structural and functional connectivity studies. Structural connectivity is typically assessed using diffusion tensor imaging (DTI) or diffusion weighted imaging (DWI), which estimates white matter connectivity patterns in the brain. Aside from the static physical organization of the brain, dynamic functional connectivity between brain regions can be assessed by magnetoencephalography (MEG), electroencephalography (EEG), and functional magnetic resonance imaging (fMRI), with only the latter allowing for the imaging of deep brain structures such as the amygdala. Functional connectivity assesses synchronized activity (e.g., changes in blood flow as indexed by blood-oxygen-level dependent, BOLD, signal fluctuations) between spatially separated brain regions. This activity can be task-related, such as when participants undergo a conditioning procedure, or spontaneous, such as when participants are told to just rest. Synchronized resting-state activity is thought to reflect 'intrinsic' coupling between brain regions, and shows high consistency across individuals (Damoiseaux et al., 2006). Notably, networks identified with resting state-fMRI closely map onto networks obtained using DWI and DTI (Collin, Sporns, Mandl, & van den Heuvel, 2014; van Oort, van Cappellen van Walsum, & Norris, 2014), suggesting that inter-regional co-activation may reflect direct anatomical links. The advantage of studying functional connectivity compared to structural connectivity is that it allows the assessment of *changes* in inter-regional coupling as a function of environmental demands.

To investigate what is known about the human extinction circuit, and compare this to animal work, we conducted a systematic literature search. Articles were identified through searches of PubMed for articles published from 1995 to 2018, using the terms "extinction" AND "fMRI" OR "imaging" OR "neuroimaging". The 571 articles resulting from this search, and relevant references cited in those articles were selectively reviewed. Articles were selected if they were peer-reviewed; they were published in English; they presented new data obtained from human participants, or a meta-analysis of such data; and fear conditioning and extinction was the primary topic of investigation. Case studies, qualitative studies and pilot studies were excluded.

3.1. Human homologue of PL-amygdala connectivity

Amygdala-anterior cingulate cortex (ACC) connections have been proposed as the human homologue of rodent PL-amygdala coupling. DTI and DWI studies showed structural connectivity between the regions as a dorsal pathway connects the amygdala with the anterior cingulate cortex (ACC) (Bracht et al., 2009; Croxson et al., 2005; Kim et al., 2011) (Fig. 1). With regard to fear-related activity *within* these regions, the role of the amygdala in human fear conditioning and extinction is not unequivocal, in contrast to the overwhelming evidence from animal studies that place the amygdala at the centre of emotional learning. Whereas early studies did report amygdala activity during fear acquisition and/or extinction (Knight, Smith, Cheng, Stein, & Helmstetter, 2004; LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998; Linnman, Zeidan, Pitman, & Milad, 2012; Milad et al., 2007; Phelps, Delgado, Nearing, & LeDoux, 2004), more recent studies have failed to replicate this. Recent meta-analyses of fMRI studies on fear conditioning (Fullana et al., 2016) and extinction (Fullana et al., 2018) failed to identify robust involvement of the amygdala in both fear acquisition and extinction (see also Section 4). In contrast, many fMRI studies have shown dACC activity in response to conditioned stimuli (Büchel, Morris, Dolan, & Friston, 1998; Cheng, Knight, Smith, Stein, & Helmstetter, 2003; Knight et al., 2004; Milad et al., 2007; Sehlmeyer et al., 2011). Meta-analyses by Fullana and colleagues and Sehlmeyer and colleagues, confirmed that dACC is robustly activated, both during fear conditioning (Fullana et al., 2016; Sehlmeyer et al., 2009) and extinction (Fullana et al., 2018). The dACC is hypothesised to receive signals from the anterior insula about the subject's cognitive, affective and physical state to facilitate homeostatic autonomic and behavioural responding (Critchley, 2009). Indeed, the meta-analysis by Fullana et al. (2018) demonstrated insula involvement in extinction learning. Co-activation of the ACC and insula is strongly associated with activation of a negative affective state such as threat anticipation (Etkin, Egner, & Kalisch, 2011; Medford & Critchley, 2010). Furthermore, dACC hyperactivity has been associated with maladaptive fear responding. For example, PTSD Patients showed heightened dACC activity during fear conditioning (Bremner et al., 2005; Rougemont-Bücking et al., 2011), and persistently enhanced dACC activity during extinction and extinction recall (Marin et al., 2016; Milad et al., 2009; Rougemont-Bücking et al., 2011). Also, individuals with a genetic vulnerability to develop anxiety pathology demonstrated enhanced ACC activity during extinction (Hermann et al., 2012). These findings were not replicated in patients with obsessive-compulsive disorder (OCD) (Milad et al., 2013), nor in individuals with high trait anxiety, where in fact the opposite (reduced involvement of the dACC during extinction) was found (Sehlmeyer et al., 2011). Furthermore, it has been suggested that sex differences might modulate ACC hyperactivity during extinction recall in PTSD (Shvil et al., 2014).

Even though there is no robust evidence for activation in the amygdala in conditioning or extinction, there are some indications that individual differences in connectivity between amygdala and ACC may berelated to fear conditioning and extinction. Functional connectivity studies showed that amygdala-dorsal ACC coupling was increased from pre- to post-conditioning (Feng, Feng, Chen, & Lei, 2014), and that amygdala-dACC coupling was altered in post-traumatic stress disorder (PTSD) (Brown et al., 2014; Sripada et al., 2012).

Together these findings have led to the proposition that dACC-amygdala projections might constitute the human homologue of rodent PL-amygdala coupling to enhance the expression of fear (Fig. 1). It is important to note that, although much overlap has been found between functional and structural connectivity (Collin et al., 2014; van Oort, van Cappellen van Walsum, & Norris, 2014), a correlation between regions does not always indicate direct anatomical connections. For example, functional connectivity has been observed between the BA and dACC, and CEA and dACC (Roy et al., 2009), whereas rodent and primate neuroanatomy shows that the cingulate area is structurally connected

to the BA, but not CEA (Freese & Amaral, 2009; Orsini & Maren, 2012). It is unclear whether connectivity between these regions could be different in the human brain compared to other primates or rodents, or whether instead, co-activation of CE and dACC may be mediated by indirect pathways. An obvious candidate is the BA, which connects to both areas.

3.2. Human homologue of IL-amygdala connectivity

Converging evidence points towards the ventromedial PFC (vmPFC) as the human homologue of rodent IL cortex (Fig. 1). Studies on both functional (Feng, Zheng, & Feng, 2016; Hare et al., 2008; Pezawas et al., 2005) and structural (for review see Kim et al., 2011; Kim & Whalen, 2009) coupling showed connectivity between the amygdala and vmPFC. Generally, adaptive emotion regulation is characterized by increased prefrontal activity and concurrently decreased amygdala activity (Delgado, Nearing, LeDoux, & Phelps, 2008; Erk et al., 2010; Hariri, Bookheimer, & Mazziotta, 2000; Hariri, Mattay, Tessitore, Fera, & Weinberger, 2003; Lieberman et al., 2007; Ochsner, Bunge, Gross, & Gabrieli, 2002; Phan et al., 2005; Urry et al., 2006). A DTI study similarly observed that strength of the connection between the amygdala and vmPFC was negatively correlated with trait anxiety (Kim & Whalen, 2009). Together, these studies suggest that strong amygdala-vmPFC connectivity is related to adaptive behavior, such as low anxiety levels in the absence of threat.

More specific to extinction, the vmPFC is generally considered the main structure involved in inhibition of conditioned responding. Several studies observed vmPFC activity during extinction learning (Gottfried & Dolan, 2004; Linnman et al., 2012; Milad et al., 2007; Phelps et al., 2004) and extinction recall (Hermann, Stark, Milad, & Merz, 2016; Kalisch et al., 2006; Linnman et al., 2012; Lonsdorf, Haaker, & Kalisch, 2014; Milad et al., 2007). The recent meta-analysis on extinction by Fullana et al. (2018) did identify a role for the vmPFC during extinction recall but not during extinction learning. It was argued that the specific contrasts used for analysing extinction learning and extinction recall could account for these unexpected results. That is, for extinction learning the typical contrast (CS+ > CS−) could result in a minimal difference given that during extinction the CS+ becomes a safety signal (similar to the CS−), and the vmPFC responds strongly to safety signals (Harrison et al., 2017; Schiller & Delgado, 2010) The vmPFC may thus respond equally strong to the CS+ and CS− stimuli, and not 'show up' in the contrast. In contrast, during extinction *recall* usually an extinguished stimulus (CS+E) is contrasted with a stimulus that was not presented during extinction and thus not extinguished (CS+U). In such a contrast (i.e., CS+E > CS+U), the extinguished stimulus will have gained safety properties and engage the vmPFC, whereas the stimulus that was not presented will not, leading to differential vmPFC activity.

Yet, while the vmPFC (see Section 3.1) may not be differentially engaged during extinction learning, individual differences in vmPFC and amygdala connectivity do seem to be associated with extinction learning. For example, resting state functional connectivity between vmPFC and amygdala was associated with extinction success (Feng et al., 2016). In individuals with a polymorphism of the BDNF gene (related to extinction plasticity, see Section 2.1), impaired extinction learning (as measured by SCR) was associated with reduced vmPFC activity during extinction (Soliman et al., 2010). In contrast, the ability to flexibly regulate emotions (i.e., cognitive reappraisal) was associated with enhanced vmPFC activity during extinction recall (Hermann, Keck, & Stark, 2014).

Patients suffering from anxiety disorders, PTSD, or OCD typically demonstrate impaired extinction recall (Blechert, Michael, Vriends, Margraf, & Wilhelm, 2007; Duits et al., 2015, 2016; Jovanovic & Norrholm, 2011; McLaughlin et al., 2015). This could be explained by hypoactivation of the vmPFC during extinction recall (Garfinkel et al., 2014; Marin et al., 2016, 2017; Milad et al., 2009, 2013; Rougemont-

Bücking et al., 2011), with more pronounced hypoactivation for those individuals with more than one anxiety disorder (Marin et al., 2017), and more pronounced hypoactivation for adults compared to adolescents with anxiety disorders (Britton et al., 2013). Furthermore, anxiety-related alterations of vmPFC-amygdala coupling during extinction recall also seem to change with age (Gold et al., 2016). Recently, vmPFC activity during extinction in patients was identified as a marker of successful exposure in patients; engagement of the vmPFC during extinction predicted exposure treatment success in patients with fear of public speaking (Ball, Knapp, Paulus, & Stein, 2017). In PTSD impaired extinction may be the result of vmPFC hypoactivity together with dACC hyperactivity. The persistently enhanced dACC activity during extinction recall (see Section 3.1) coincided with reduced vmPFC activity (Milad et al., 2009). Thus, in individuals with PTSD, regions responsible for safety learning and emotion regulation failed to participate, whereas regions responsible for fear expression remained engaged during extinction learning.

3.3. Hippocampus

Similar to rodents, the hippocampus has been ascribed a central role in extinction in humans (Åhs, Kragel, Zielinski, Brady, & LaBar, 2015; Hermann et al., 2016; Hermann, Stark, Blecker, Milad, & Merz, 2017; Kalisch et al., 2006; Lissek, Glaubitz, Uengoer, & Tegenthoff, 2013; Milad et al., 2007, 2009; Visser, Kunze, Westhoff, Scholte, & Kindt, 2015; Visser, Scholte, Beemsterboer, & Kindt, 2013). In PTSD patients, impaired extinction recall is associated with hippocampal hypoactivation (Garfinkel et al., 2014; Marin et al., 2016; Milad et al., 2009; Rougemont-Bücking et al., 2011). Recent studies suggested that co-activation and connectivity patterns of the hippocampus to other regions involved in extinction learning (amygdala, ACC, vmPFC) may be similar to that in rodents (Hermann et al., 2017, 2016); the left hippocampus and left insula showed stronger responses to a CS+ compared to a CS− one day after the CS+ was extinguished (i.e., extinction recall), while the right amygdala and right hippocampus showed reduced responses. Additionally, individuals who showed stronger extinction recall (as indexed by SCR difference scores between CS+ and CS) showed more activation in the vmPFC and left hippocampus. Functional connectivity between these regions was not reported for extinction recall, but seemed enhanced in individuals who, after a subsequent shift to an unfamiliar context, showed stronger renewal of conditioned fear (SCRs). A concurrent DTI study failed to find clear evidence of structural alterations underlying these individual differences (Hermann et al., 2017). Thus, while it can be assumed that, analogous to the rodent network, the hippocampus projects to both prefrontal areas and the amygdala (Fig. 1), further research should reveal the subregions involved in these projections.

It should be noted that while in animals the hippocampus is strongly linked to contextual modulation, the neural mechanisms of contextual processing are difficult to examine in humans. The options for establishing different contexts inside an MRI scanner are limited (Maren et al., 2013). Context manipulations in neuroimaging studies are therefore restricted to changing visual or auditory backgrounds of the stimulus material, and participants cannot freely move in these contexts. The development of virtual-reality technologies, where contexts can be more immersive, and the participant may experience to some extent experience free movement may be promising in this regard. Also, while in animals a clear dual role (i.e., modulating both activation and inhibition of fear expression) has been identified for the hippocampus, this is less clear in humans. The dual role might obscure hippocampal involvement in imaging studies; a meta-analysis on extinction (Fullana et al., 2018) found robust involvement of the hippocampus during extinction recall, but not extinction learning, in the same way that vmPFC activation is only observed during extinction recall: newly acquired safety information for the CS+ may obscure differences between CS+ and CS−. A first step could be to replicate the animal findings that

hippocampal inactivation during extinction and extinction recall differentially affect fear expression. Direct manipulation of the hippocampus is of course impossible in humans. However, an animal study demonstrated that higher doses of systemic administration of a scopolamine, a cholinergic antagonist (known to affect the hippocampus) before extinction severely impaired extinction learning and recall (Zelikowsky et al., 2013). It would be exciting to see whether similar effects could be observed in humans, and to extend the findings by investigating whether scopolamine before extinction recall enhances extinction generalization.

Finally, extinction after trace conditioning (i.e., where the UCS follows CS offset after a delay), is a relatively understudied topic, and our review identified a gap in the literature when it comes to combining this paradigm with neuroimaging. Given that conditioning paradigms have revealed a far more robust involvement of the hippocampus during the acquisition of fear in studies employing trace conditioning compared to delay conditioning (Sehlmeyer et al., 2009), combining extinction of trace-conditioned fear with neuroimaging might be a promising next step for gaining insight into the role of hippocampus.

3.4. Pharmacological manipulation of extinction

In animals, highly specific manipulations via optogenetic activation/inhibition or local infusion of pharmacological agent can be performed. These manipulations contribute to the characterization of the fear extinction circuitry. For example, pharmacological inactivation of the hippocampus relieves extinction of its contextual dependency (Corcoran et al., 2005; Corcoran & Maren, 2001) and, thus, points towards a role for the hippocampus in contextual processing during extinction and its recall (Section 2.2). While pharmacological agents cannot be administered locally in humans, these studies may, nevertheless, provide invaluable information about the neural circuitry underlying extinction, and could be a crucial translational step in the science-driven optimization of treatments for anxiety disorders. Various lines of research have addressed the effect of pharmacological agents on extinction.

First, research suggests that the stress hormone cortisol may impair the recall of declarative emotional memories, while enhancing their consolidation (Schwabe, Joëls, Roozendaal, Wolf, & Oitzl, 2012). Translating these findings to the conditioning paradigm, a multi-day study tested whether administration of cortisol before the start of extinction would enhance extinction memory (tested a week later) and would facilitate the transfer of the extinction memory from the extinction context to a new context (Merz, Hamacher-Dang, Stark, Wolf, & Hermann, 2018). Results showed that cortisol administration before extinction training diminished conditioned fear responses during extinction learning. This was evident in reduced differential skin conductance levels, and reduced differential neural activation at the beginning of extinction in the amygdala, hippocampus, and anterior parahippocampal gyrus. In contrast, the hippocampus showed enhanced CS+/CS− differentiation activation, and enhanced vmPFC coupling, at the beginning of extinction recall. These findings are in line with the idea that cortisol strengthens the consolidation of extinction memory leading to stronger extinction recall. This is possibly mediated by synchronised action of the hippocampus and vmPFC, inhibiting conditioned fear responses via projections to inhibitory intercalated cells in the amygdala. However, effects of cortisol did not transfer to a different context. Cortisol-induced enhancement of extinction learning could only be observed when a consolidation period between acquisition and extinction is taken into account. That is, cortisol administration immediately after acquisition and before extinction learning impaired extinction learning (Merz, Hermann, Stark, & Wolf, 2014). Also, it is important to note that many of these studies have assessed cortisol effects in either all-male (Hermann, et al., 2014; Merz et al., 2018; Merz, Hamacher-Dang, & Wolf, 2014) or all-female (Tabbert et al., 2010) populations, even though (or because) sex differences play an

important role in the modulation of fear and extinction learning (Merz et al., 2012), and cortisol may exert gender-specific effects on extinction learning. For example, cortisol administration prior to tests of return of fear increased fear renewal following reinstatement (but not renewal only) in men, but not in women. In men, this coincided with enhanced activation in the right amygdala in response (CS+ > CS−), while in women activation in this region decreased after cortisol treatment (Kinner, Wolf, & Merz, 2018).

Second, noradrenaline is involved in the consolidation of extinction learning (Mueller & Cahill, 2010) and animal studies revealed that noradrenergic receptors are abundant in areas related to fear and extinction learning, including the amygdala, hippocampus and the vmPFC (Kim & Jung, 2006; LeDoux, 2000; Milad & Quirk, 2012). Hence, enhancing noradrenaline was considered a potential candidate for boosting extinction learning. However, administration of the selective NA reuptake inhibitor reboxetine (RBX), in animals known to enhance noradrenergic levels in frontal and hippocampal areas (Hajós, Fleishaker, Filipiak-Reisner, Brown, & Wong, 2004), following extinction affected neither behavioural (SCR), nor neural correlates of extinction learning and recall (Lonsdorf, Haaker, Fadai, & Kalisch, 2014). Another study found that noradrenergic blockade even boosted extinction learning (Kroes et al., 2016). Administration of the beta-blocker propranolol prevented spontaneous recovery and reinstatement of SCR (but not subjective fear). Also, hippocampal activity during extinction recall was increased in those receiving propranolol combined with extinction. The finding that increased hippocampal activity during extinction recall is associated with reduced return of fear and thus enhanced extinction plasticity is in line with recent human data (Fullana et al., 2018; Hermann et al., 2017, 2016).

Third, another class of receptors abundant in the supposed extinction circuitry is the cannabinoid receptor (Davies, Pertwee, & Riedel, 2002; Mackie, 2005; Wilson & Nicoll, 2002). Infusion of a cannabinoid receptor antagonist in the hippocampus or IL blocks extinction consolidation, while activation of these receptors results in enhanced extinction learning and recall (Bitencourt, Pamplona, & Takahashi, 2008; de, Pasqualini, Diehl, Molina, & Quillfeldt, 2008; Lin, Mao, Su, & Gean, 2009). A first human imaging study combining a cannabinoid agonist (tetrahydrocannabinol (THC)) with extinction did not observe enhanced extinction learning or recall (SCR). There was increased activity in the vmPFC and hippocampus during extinction recall. These findings are hard to interpret given that the placebo group, unexpectedly, showed no involvement of these areas during recall and the lack of behavioural evidence of extinction enhancement. Note that the effects of dopamine administration on extinction in humans are described in Section 3.5.2. Together, more imaging studies on pharmacological enhancement of extinction are needed to explain these mixed results.

3.5. Prediction error and extinction in humans

Prediction error is the driving force of extinction learning and therefore a possible route to enhance extinction learning is to maximize prediction error (Craske, 2015; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; Culver, Vervliet, & Craske, 2015). Promisingly, this approach is taking ground in clinical practice. Traditionally, clinicians were trained to consider initial fear elevation followed by fear reduction within the same session as a sign of therapy efficacy. However, as discussed above, within-session extinction is a poor predictor of longer-term reduction in fear responses, that is, across sessions. Instead of focusing on immediate fear elevation, present exposure therapies are rather designed such that the experience maximizes violation of expectancies (Craske et al., 2014; Craske, 2015). The neural circuitry underlying extinction-driven PE and in humans is however scarcely investigated.

3.5.1. Hippocampus and PE in humans

fMRI studies revealed activation in response to novel stimuli in a

network that included the hippocampus (for review see Kafkas & Montaldi, 2018). Furthermore, in different memory paradigms (i.e., implicit and explicit associative memory recognition tasks) PE-related hippocampal activity has been observed (Duncan, Curtis, & Davachi, 2009; Kumaran & Maguire, 2006; Long, Lee, & Kuhl, 2016). High resolution (hr)-fMRI studies identified, similar to the animal findings, a key role for PE signaling in the CA1 area of the hippocampus (Chen, Cook, & Wagner, 2015; Chen, Olsen, Preston, Glover, & Wagner, 2011; Duncan, Ketz, Inati, & Davachi, 2012). Imaging studies on PE during extinction are lacking (but see one study on PE-related activity during extinction recall; Spoormaker et al., 2012). Yet, hippocampal activity in response to unreinforced CS presentations has been proposed to be the neural correlate of US omission or a negative PE (Fullana et al., 2016; but see Spoormaker et al., 2011).

While we await further research to reveal the neurobiological correlates of PE, it should be noted that in human conditioning a behavioural index of PE is available, as demonstrated in studies on memory reconsolidation. Reconsolidation is the phenomenon that memories can regain plasticity after consolidation. Animal studies have shown that, upon its retrieval, a memory may again become dependent on protein synthesis, and protein synthesis interfering agents can block the memory restabilization, resulting in fear amnesia (Nader, Schafe, & LeDoux, 2000; Sara, 2000). While protein synthesis inhibitors are not safe to use in humans, pharmacological procedures involving administration of the β-noradrenergic antagonist propranolol shortly after re-exposure to the feared stimulus have also been shown to disrupt reconsolidation processes, resulting in a persistent reduction of the fear response (Kindt, Soeter, & Vervliet, 2009; Soeter & Kindt, 2010, 2011). Typically, reconsolidation is induced by short re-exposure to the feared stimulus, while prolonged re-exposure engages extinction (Bustos, Maldonado, & Molina, 2008; Lee, Milton, & Everitt, 2006). PE is crucial in both reconsolidation (memory updating) and extinction (new learning), yet the frequency and duration of exposure to the feared stimulus determines which of these processes is initiated. From a clinical perspective, it is crucial to establish markers to indicate when a PE occurs. This is not only relevant in the context of maximizing PE as a route to enhance the effectiveness of exposure therapy (Craske, 2015; Craske et al., 2014; Culver et al., 2015), but also to demarcate the transition from reconsolidation to extinction processes. This is particularly important when the aim is to pharmacologically disrupt or boost one of these processes: one may inadvertently enhance the original fear memory via reconsolidation update mechanisms (e.g., using NMDA agonists such as D-cycloserine), or hamper extinction learning (e.g., using noradrenergic antagonist such as propranolol). The great advantage of studying humans in the context of fear learning is their ability to verbalize the relationship between the CS and US. In studies on memory reconsolidation, a single PE (indexed by changes in US-expectancy ratings) served as an independent index of whether memory reconsolidation was or was not induced (Sevenster, Beckers, & Kindt, 2013, 2014). There was some evidence that two PE's (a reduction in US-expectancy during the first trials of extinction learning) corresponded to the transitional phase between reconsolidation and extinction (Sevenster, Beckers, & Kindt, 2014). Hence, the US-expectancy ratings served as an index for PE, to such an extent that it could differentiate between different memory processes engaged by re-exposure to a feared stimulus. Thus far, studies have not identified the point at which extinction is engaged during re-exposure to the CS. It should be noted that the number or duration of trials that corresponds to a certain underlying memory process depends on the learning history. That is, the number of re-exposure trials that engage extinction will differ following a short fear acquisition session relative to a long acquisition session (Sevenster et al., 2013, 2014). Future imaging studies on the neural correlates of extinction may use a behavioural index of PE to help identify and focus on these re-exposure trials that trigger PE.

3.5.2. Prediction error and the ventral tegmental area (VTA) in humans

Human imaging data concurs with the idea that the hippocampal-VTA loop plays an important role in reward-based memory encoding. Activation of the loop was associated with successful formation of reward memory (Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Callan & Schweighofer, 2008; Gruber, Gelman, & Ranganath, 2014; Kahn & Shohamy, 2013; Wittmann et al., 2005; Wolosin, Zeithamova, & Preston, 2012). However, to our knowledge, there are no studies on the role of hippocampal-VTA connection in extinction in humans. Given that dopamine projections from the VTA to the hippocampus are implicated in memory encoding (see Section 2.3.2), we will focus on the effects of pharmacological manipulation of dopamine on extinction.

There is preliminary evidence that manipulating dopaminergic activity can have beneficial consequences for extinction learning in humans. A translational study first showed that post-extinction administration of L-dopa – a chemical precursor that enhances brain dopamine levels – improved extinction in mice, evidenced by reduced spontaneous recovery, reinstatement and renewal of fear (Haaker et al., 2013). Promisingly, the animal findings were translated to humans in follow-up studies. A single dose of L-dopa following extinction prevented renewal of conditioned skin conductance response (SCR) (Haaker et al., 2013; but see Haaker, Lonsdorf, & Kalisch, 2015). The participants receiving placebo did demonstrate renewal, associated with vmPFC deactivation, whereas this deactivation was not observed in the L-dopa group. Instead, enhanced vmPFC activity during renewal was linked to decreased amygdala activity (Haaker et al., 2013, 2015). The participants showed generalization of the extinction training to other contexts. The finding that dopamine-enhancing agents prevented renewal of fear responses and facilitated extinction learning is promising for the development of (dopamine) pharmacotherapy in the treatment of anxiety disorders. The findings indicate that dopamine could relieve extinction of its context-dependency. Indeed, a first clinical trial showed that the dopamine enhancing agent Methylenedioxymethamphetamine (MDMA) resulted in long lasting enhancement of therapy efficacy in treatment-resistant PTSD patients (Mithoefer, Wagner, Mithoefer, Jerome, & Doblin, 2011; Mithoefer et al., 2013). Notably, it remains unclear whether effects of L-dopa and MDMA are solely attributable to their effects on dopamine since these agents also prominently affect serotonin (Farré et al., 2007; Liechti & Vollenweider, 2001).

While it is encouraging that dopamine administration may facilitate extinction learning, it is important to note that several studies applied dopamine *after* extinction learning. Hence, these effects cannot be attributed to online dopamine-facilitated negative PE during extinction learning. Recent animal studies showed that natural enhancement of dopamine levels by novelty manipulation that preceded extinction facilitated both extinction learning and retention (de Carvalho Myskiw, Furini, Benetti, & Izquierdo, 2014; Menezes et al., 2015). Ideally, a noninvasive method for enhancing extinction learning is preferred over drug administration. Novelty manipulations aimed at increasing dopamine levels might be a promising new avenue for therapy improvement.

4. Challenges and future directions in mapping fear and extinction networks in humans

Most of our knowledge about extinction learning and its underlying mechanisms comes from work in non-human animals. In this review we have aimed to compare this work to what cognitive neuroscience has taught us about this process in humans. In this section we discuss the problems that arise when drawing parallels between, what turns out to be, fundamentally different levels of observation. To illustrate this, we start with what is probably the most undisputed centre of associative fear memory in animals: the amygdala.

Although the amygdala is generally regarded as the brain's integrative centre for emotions, and critical for emotional learning (LeDoux, 2003), its role in human associative fear learning and

extinction is, as is evident from Section 3.1, not clear. Recent, comprehensive meta-analyses on human fear conditioning (Fullana et al., 2018, 2016) failed to find robust amygdala activation during acquisition and extinction of fear, while an earlier influential meta-analysis (Sehlmeyer et al., 2009) reported amygdala activation in about half of the included studies. It has often been argued that amygdala responsiveness habituates rapidly (Büchel et al., 1998; LaBar et al., 1998; Lindner et al., 2015), and activation may therefore be obscured when analysing over multiple trials (i.e., the entire fear conditioning epoch). However, the same meta-analysis (Fullana et al., 2016) showed that, even when acquisition was divided into an early and late phase, no evidence was found for transient amygdala activity.

It is also argued that absence of amygdala activity in some reports might be due to differences in study design (Sehlmeyer et al., 2009). For example, some studies have used fear-relevant stimuli, such as spiders, as conditioned stimuli (van Well, Visser, Scholte, & Kindt, 2012; Visser, de Haan, et al., 2016), whereas others used simple geometric shapes (Jensen et al., 2003), or neutral faces and houses (Visser, de Haan, et al., 2016; Visser et al., 2015, 2013; Visser, Scholte, & Kindt, 2011). However, so far, meta-analyses have not been able to find a systematic pattern between studies that do find amygdala activation, and the studies that do not: Activation of the amygdala was independent of modality (visual, acoustic, olfactory) of either CS or UCS, and independent of reinforcement rate (partial vs. full reinforcement) (Sehlmeyer et al., 2009). Also, no clear picture emerges when comparing studies that used fear-relevant stimuli as CS (e.g., neutral faces, or pictures of spiders) compared to studies that used neutral stimuli (geometric shapes used in the different studies (Fullana et al., 2016; Sehlmeyer et al., 2009).

An alternative explanation for the many null-findings is that fMRI may not be suited for imaging the amygdala, with its different subnuclei that putatively have dissociable or even opposite functions. Rodent studies identified an intricate network consisting of reciprocal cortical-amygdaloid and inter-amygdala pathways. On top of that, some nuclei showed involvement in both excitatory and inhibitory processes of fear and safety learning, respectively (Orsini & Maren, 2012; Quirk & Mueller, 2008). The relatively low spatial resolution of fMRI (as opposed to, for example, single-cell recording) does not permit the mapping of such a small-scale organization. Apart from its low spatial resolution, fMRI suffers from signal dropouts and image distortions, specifically in structures located deep within the temporal lobe (Fullana et al., 2016; Mechias, Etkin, & Kalisch, 2010; Sehlmeyer et al., 2009). At the same time, false positive findings may arise as a result of the amygdala's proximity to large veins, which is especially problematic for emotion research, where (non-specific) changes in blood flow co-vary with stimulus properties (Boubela et al., 2015). Thus, we may simply lack the tools to image reliably the amygdala in humans. Decades of animal work render it highly plausible that the amygdala plays a key role in associative fear learning, but Fullana et al. (2016, 2018) rightfully point out that we should acknowledge that, thus far, fear conditioning and extinction studies have failed to show consistent responding in the amygdala circuitry.

New tools to investigate neural processing during fear and extinction learning are required to shed light on the issue. Traditionally, fMRI analyses have focused on mapping activity in individual brain voxels or regions during learning tasks. In contrast, multi-voxel pattern analysis (MVPA) aims to identify responses across groups of voxels, to characterize the unique neural representation of a stimulus within a certain brain region (Haxby et al., 2001). Thus, whereas conventional methods compared a voxel's or region's signal strength between conditions, MVPA recognizes the unique contribution of multiple voxels within a population. Accordingly, a voxel that would not pass the test when considered separately can be identified with MVPA to make a significant contribution within a pattern of responses. By comparing or classifying patterns related to different stimuli, one can assess the degree to which different stimuli or cognitive states are alike (Haynes &

Rees, 2005; Kamitani & Tong, 2005; Kriegeskorte, Mur, & Bandettini, 2008; Norman, Polyn, Detre, & Haxby, 2006). Techniques that evaluate unique patterns of responses, rather than average signal strength, offer increased sensitivity by capitalizing on regional variations (possibly sub-nuclei), rather than masking them (for review see Davis & Poldrack, 2013; Norman et al., 2006). In fear conditioning, this enhanced sensitivity is exemplified by a series of studies that did not reveal amygdala responding during fear acquisition or extinction when performing traditional activation-based analyses (Visser, de Haan, et al., 2016; Visser et al., 2015, 2013. 2011). However, fear learning was evident from trial-to-trial changes in neural *patterns* of responses: stimuli that were paired with a shock, but not stimuli that were never paired with a shock, showed an increase in similarity of activity patterns in the amygdala over the course of conditioning (Visser et al., 2015, 2013). Similar changes in neural activation patterns as a result of fear conditioning have been reported in other studies (Bach, Weiskopf, & Dolan, 2011; Braem et al., 2017; Dunsmoor, Kragel, Martin, & LaBar, 2014; Li, Howard, Parrish, & Gottfried, 2008). Responses identified with pattern analysis were by no means restricted to the amygdala, or even largest in the amygdala, but did suggest that the amygdala is indeed involved in associative learning, along with numerous other regions.

This highlights another challenge: Human neuroimaging research frequently reveals neural responses in areas other than those typically studies in animals (e.g., visual association areas, Fullana et al., 2018). The use of pattern analysis techniques additionally reveals effects outside the traditional human fear circuit (e.g., in the superior frontal gyrus (e.g., in the superior frontal gyrus; Visser, de Haan, et al., 2016; Visser et al., 2015, 2013, 2011; the piriform cortex; Li et al., 2008). Similarly, evidence for extinction of fear, as indexed by a decrease in differential pattern similarity, was not only observed in the amygdala, hippocampus and vmPFC but in numerous other areas (e.g., the superior frontal gyrus; occipitotemporal regions) outside the traditional extinction circuitry (Visser, de Haan, et al., 2016; Visser et al., 2015, 2013, 2011). Such findings pose additional challenges to bridging different levels of research. Note that finding effects in 'non-traditional' areas in itself does not necessarily indicate a discrepancy between human and rodent circuits (or a discrepancy between humans in 'pattern analysis' studies and in humans in previous neuroimaging studies for that matter). In most animal studies, where a priori hypotheses guide the selection of cells from which to record, effects that occur in other, less typical areas will most likely go unnoticed. This does not implicate that there are no neurons in other regions involved in the processes of interest. As another example, recent studies showed that the cerebellum contributes to fear extinction learning (Chang et al., 2015; Utz et al., 2015), an observation that was confirmed by a recent meta-analysis of extinction (Fullana et al., 2018). The cerebellum is however often ignored, or in many human neuroimaging research not even scanned. Also, while individual studies neglected involvement of the dorsolateral prefrontal cortex (dlPFC), it has been suggested that the dlPFC plays a prominent role in extinction learning and recall (Fullana et al., 2018). The dlPFC is implicated in cognitive control and emotion regulation (Hartley & Phelps, 2010; Miller & Cohen, 2001). It is likely that extinction is at least partly affected by higher order cognitive strategies to regulate emotion (Schiller & Delgado, 2010).

What these findings demonstrate is that whether one can detect an effect in a particular area depends on where data are collected and what method is employed, with for example neural pattern analysis being a more sensitive technique for detecting learning-dependent changes than analysis of average activation (Bach et al., 2011; Visser, de Haan, et al., 2016; Visser et al., 2015, 2013, 2011). Stepping away from traditional activation-based analyses may also provide a solution to some of the problems associated with imaging subcortical structures in humans. Perhaps one day new analysis methods such as MVPA may help identify subregions in the fear and extinction circuitry and advance our understanding of the complex interplay between structures within the neural circuitry underlying associative learning, especially when

combined with high-resolution functional MRI (hr-fMRI), and/or high field-strength imaging, allowing for an even finer-grained mapping of regional variations (Balderston, Schultz, Hopkins, & Helmstetter, 2015). Yet, given the wealth of rodent data showing within-region intermixing of neuron types that respond in different and even opposite ways to conditioned and extinguished cues, it seems highly unlikely that even ultra high-resolution BOLD-MRI will ever be able to inform about processes at the scale of neural populations. Moreover, such imaging will always be restricted by the spatial and temporal resolution of the BOLD signal, which is much lower than that of neuronal signalling.

The question is whether comparisons across species at such microscopic levels are really necessary, or whether we can learn from similarities in organisational structures at a larger scale (e.g., Kriegeskorte et al., 2008). Aside from examining similarities in spatial organisation, some elements of extinction learning, such as trial-to-trial changes as a function of different manipulations can be validly investigated using fMRI, when combined with MVPA (the high sensitivity of MVPA compared to analysis of average activation allows for single-trial analysis; e.g., Visser, de Haan, et al., 2016). In addition, MVPA is well suited to assess similarities in the neural representations of stimuli and therefore ideal to study the generalization of fear (Dunsmoor et al., 2014; Visser, Haver, Zwitser, Scholte, & Kindt, 2016; Visser et al., 2015) and potentially generalisation of extinction. While this does not in itself reveal insights into the degree to which human and animal circuits are comparable, such techniques do provide another read-out for learning and memory processes that can be used to bridge findings from animals to humans (Visser et al., 2018).

5. Conclusion

In this review, we investigated the degree to which state of the art research has provided evidence for differences and similarities between the neural circuitry underlying extinction learning in non-human animals and humans. At a macroscale, networks in rodents and humans seem to overlap to some extent (e.g. PL and ACC; IL and vmPFC in rodents and humans respectively), but current imaging techniques preclude comparisons at a smaller scale, especially in areas that are functionally heterogeneous, such as the amygdala. Moreover, human neuroimaging shows the involvement of numerous areas (insula, visual association areas, cerebellum) in human fear extinction, areas that are not typically studied in animals. Findings show that effects are partly dependent on the methods employed, not only across species, but also across human neuroimaging studies.

There is currently little knowledge of extinction-related PE in humans. It is surprising that the neural correlates of PE in animals also constitute a relatively understudied area. The hippocampal-VTA loop has been proposed to control memory encoding (Lisman & Grace, 2005) and may therefore play an important part in mediating PE during extinction learning. Thus far, this remains mainly a theoretical construct.

Animal studies have significantly advanced the identification of the physical substrate of a memory, the memory engram (Josselyn et al., 2015; Tonegawa et al., 2015), with highly advanced techniques, such as optogenetics, allowing for the mapping of microcircuits involved in extinction learning and recall, and the potential to better characterise the role of PE in these processes. Human research is, however, lagging behind, given that the study of memory representations relies on relatively coarse neuroimaging techniques such as fMRI. New techniques such as multi-voxel pattern analysis and high-resolution imaging have opened up new possibilities to study processes involved in extinction at a smaller scale. Furthermore, the higher sensitivity of multi-voxel analysis techniques compared to traditional approaches allows for single-trial analysis of neural activation patterns, and thus a more fine-grained study of the temporal dynamics of extinction learning, as well as generalisation of fear (or safety) to similar stimuli. However, it seems unlikely that even ultra-high-resolution BOLD-MRI will ever be able to

inform about processes at the scale of neural populations, not in the least because such imaging will always be restricted by the spatial and temporal resolution of the BOLD signal, which is much lower than that of neuronal signalling. Translational progress will therefore continue to depend on a variety of techniques and clever behavioural and pharmacological manipulations, which – when combined – may more indirectly allow inferences regarding the underlying neurobiology.

Conflict of interest

De authors declare no conflict of interest.

References

Abraham, A. D., Neve, K. A., & Lattal, K. M. (2014). Dopamine and extinction: A convergence of theory with fear and reward circuitry. *Neurobiology of Learning and Memory, 108*, 65–77.

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. E. (2006). Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron, 50*, 507–517.

Åhs, F., Kragel, P. A., Zielinski, D. J., Brady, R., & LaBar, K. S. (2015). Medial prefrontal pathways for the contextual regulation of extinguished fear in humans. *NeuroImage, 122*, 262–271.

Alvarez, R. P., Johnson, L., & Grillon, C. (2007). Contextual-specificity of short-delay extinction in humans: Renewal of fear-potentiated startle in a virtual environment. *Learning & Memory, 14*, 247–253.

Amano, T., Unal, C. T., & Paré, D. (2010). Synaptic correlates of fear extinction in the amygdala. *Nature Neuroscience, 13*, 489–494.

Bach, D. R., Weiskopf, N., & Dolan, R. J. (2011). A stable sparse fear memory trace in human amygdala. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31*, 9383–9389.

Balderston, N. L., Schultz, D. H., Hopkins, L., & Helmstetter, F. J. (2015). Functionally distinct amygdala subregions identified using DTI and high-resolution fMRI. *Social Cognitive and Affective Neuroscience, 10*, 1615–1622.

Ball, T. M., Knapp, S. E., Paulus, M. P., & Stein, M. B. (2017). Brain activation during fear extinction predicts exposure success. *Depression and Anxiety, 34*, 257–266.

Barad, M. (2005). Fear extinction in rodents: Basic insight to clinical promise. *Current Opinion in Neurobiology, 15*, 710–715.

Barbeau, E. J., Chauvel, P., Moulin, C. J. A., Regis, J., & Liégeois-Chauvel, C. (2017). Hippocampus duality: Memory and novelty detection are subserved by distinct mechanisms. *Hippocampus, 27*, 405–416.

Bitencourt, R. M., Pamplona, F. A., & Takahashi, R. N. (2008). Facilitation of contextual fear memory extinction and anti-anxiogenic effects of AM404 and cannabidiol in conditioned rats. *European Neuropsychopharmacology, 18*, 849–859.

Blechert, J., Michael, T., Vriends, N., Margraf, J., & Wilhelm, F. H. (2007). Fear conditioning in posttraumatic stress disorder: Evidence for delayed extinction of autonomic, experiential, and behavioural responses. *Behaviour Research and Therapy, 45*, 2019–2033.

Boubela, R. N., Kalcher, K., Huf, W., Seidel, E.-M., Derntl, B., Pezawas, L., ... Moser, E. (2015). fMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. *Scientific Reports, 5*, 10499.

Bouchet, C. A., Miner, M. A., Loetz, E. C., Rosberg, A. J., Hake, H. S., Farmer, C. E., ... Greenwood, B. N. (2018). Activation of nigrostriatal dopamine neurons during fear extinction prevents the renewal of fear. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 43*, 665–672.

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

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

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

Bracht, T., Tüscher, O., Schnell, S., Kreher, B., Rüsch, N., Glauche, V., ... Saur, D. (2009). Extraction of prefronto-amygdalar pathways by combining probability maps. *Psychiatry Research, 174*, 217–222.

Braem, S., De, J. H., Demanet, J., Yuen, K. S. L., Kalisch, R., & Brass, M. (2017). Pattern analyses reveal separate experience-based fear memories in the human right amygdala. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 37*, 8116–8130.

Bramham, C. R., & Messaoudi, E. (2005). BDNF function in adult synaptic plasticity: The synaptic consolidation hypothesis. *Progress in Neurobiology, 76*, 99–125.

Bremner, J. D., Vermetten, E., Schmahl, C., Vaccarino, V., Vythilingam, M., Afzal, N., ... Charney, D. S. (2005). Positron emission tomographic imaging of neural correlates of a fear acquisition and extinction paradigm in women with childhood sexual-abuse-related post-traumatic stress disorder. *Psychological Medicine, 35*, 791–806.

Britton, J. C., Grillon, C., Lissek, S., Norcross, M. A., Szuhany, K. L., Chen, G., ... Pine, D. S. (2013). Response to learned threat: An FMRI study in adolescent and adult anxiety. *The American Journal of Psychiatry, 170*, 1195–1204.

Brown, V. M., LaBar, K. S., Haswell, C. C., Gold, A. L., McCarthy, G., & Morey, R. A. (2014). Altered resting-state functional connectivity of basolateral and centromedial amygdala complexes in posttraumatic stress disorder. *Neuropsychopharmacology, 39*, 361–369.

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

Bustos, S. G., Maldonado, H., & Molina, V. A. (2008). Disruptive effect of midazolam on fear memory reconsolidation: Decisive influence of reactivation time span and memory age. *Neuropsychopharmacology, 34*, 446–457.

Callan, D. E., & Schweighofer, N. (2008). Positive and negative modulation of word learning by reward anticipation. *Human Brain Mapping, 29*, 237–249.

Chang, D.-I., Lissek, S., Ernst, T. M., Thürling, M., Uengoer, M., Tegenthoff, M., ... Timmann, D. (2015). Cerebellar contribution to context processing in extinction learning and recall. *Cerebellum (London, England), 14*, 670–676.

Chen, J., Cook, P. A., & Wagner, A. D. (2015). Prediction strength modulates responses in human area CA1 to sequence violations. *Journal of Neurophysiology, 114*, 1227–1238.

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. *Learning & Memory, 18*, 523–528.

Cheng, D. T., Knight, D. C., Smith, C. N., Stein, E. A., & Helmstetter, F. J. (2003). Functional MRI of human amygdala activity during Pavlovian fear conditioning: Stimulus processing versus response expression. *Behavioral Neuroscience, 117*, 3–10.

Collin, G., Sporns, O., Mandl, R. C. W., & van den Heuvel, M. P. (2014). Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex. *Cerebral Cortex (New York, NY), 24*, 2258–2267.

Corcoran, K. A., Desmond, T. J., Frey, K. A., & Maren, S. (2005). Hippocampal inactivation disrupts the acquisition and contextual encoding of fear extinction. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 25*, 8978–8987.

Corcoran, K. A., & Maren, S. (2001). Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. *Journal of Neuroscience, 21*, 1720–1726.

Craske, M. G. (1999). *Anxiety disorders: Psychological approaches to theory and treatment, Vol. xxi.* Boulder, CO, US: Westview Press.

Craske, M. G. (2015). Optimizing exposure therapy for anxiety disorders: An inhibitory learning and inhibitory regulation approach. *Verhaltenstherapie, 25*, 134–143.

Craske, M. G., Treanor, M., Conway, C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10–23.

Critchley, H. D. (2009). Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicants. *International Journal of Psychophysiology, 73*, 88–94.

Croxson, P. L., Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Pinsk, M. A., Gross, C. G., ... Rushworth, M. F. S. (2005). Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 25*, 8854–8866.

Culver, N. C., Vervliet, B., & Craske, M. G. (2015). Compound extinction. *Clinical Psychological Science, 3*, 335–348.

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 13848–13853.

Davies, S. N., Pertwee, R. G., & Riedel, G. (2002). Functions of cannabinoid receptors in the hippocampus. *Neuropharmacology, 42*, 993–1007.

Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: Practices and pitfalls. *Annals of the New York Academy of Sciences, 1296*, 108–134.

Dawson, M. E., Schell, A. M., & Banis, H. T. (1986). Greater resistance to extinction of electrodermal responses conditioned to potentially phobic CSs: A noncognitive process? *Psychophysiology, 23*, 552–561.

de Carvalho Myskiw, J., Furini, C. R. G., Benetti, F., & Izquierdo, I. (2014). Hippocampal molecular mechanisms involved in the enhancement of fear extinction caused by exposure to novelty. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 4572–4577.

de, O. A., Pasqualini, G., Diehl, F., Molina, V. A., & Quillfeldt, J. A. (2008). Opposite action of hippocampal CB1 receptors in memory reconsolidation and extinction. *Neuroscience, 154*, 1648–1655.

Delamater, A. R., & Westbrook, R. F. (2014). Psychological and neural mechanisms of experimental extinction: A selective review. *Neurobiology of Learning and Memory, 38*–51.

Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron, 59*, 829–838.

Desmedt, A., Marighetto, A., & Piazza, P.-V. (2015). Abnormal fear memory as a model for posttraumatic stress disorder. *Biological Psychiatry, 78*, 290–297.

Duits, P., Cath, D. C., Heitland, I., & Baas, J. M. P. (2016). High Current anxiety symptoms, but not a past anxiety disorder diagnosis, are associated with impaired fear extinction. *Frontiers in Psychology, 7*, 252.

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., ... Baas, J. M. P. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety, 32*, 239–253.

Duncan, K., Curtis, C., & Davachi, L. (2009). Distinct memory signatures in the hippocampus: Intentional states distinguish match and mismatch enhancement signals. *Journal of Neuroscience, 29*, 131–139.

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus, 22*, 389–398.

Dunsmoor, J. E., Kragel, P. A., Martin, A., & LaBar, K. S. (2014). Aversive learning modulates cortical representations of object categories. *Cerebral Cortex (New York, NY), 24*, 2859–2872.

Eddy, K. T., Dutra, L., Bradley, R., & Westen, D. (2004). A multidimensional meta-analysis of psychotherapy and pharmacotherapy for obsessive-compulsive disorder. *Clinical Psychology Review, 24*, 1011–1030.

Eisenhardt, D. (2014). Molecular mechanisms underlying formation of long-term reward memories and extinction memories in the honeybee (*Apis mellifera*). *Learning & Memory, 21*, 534–542.

El-Ghundi, M., O'Dowd, B. F., & George, S. R. (2001). Prolonged fear responses in mice lacking dopamine D1 receptor. *Brain Research, 892*, 86–93.

Erk, S., Mikschl, A., Stier, S., Ciaramidaro, A., Gapp, V., Weber, B., & Walter, H. (2010). Acute and sustained effects of cognitive emotion regulation in major depression. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*, 15726–15734.

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences, 15*, 85–93.

Farré, M., Abanades, S., Roset, P. N., Peiró, A. M., Torrens, M., O'Mathúna, B., ... de la Torre, R. (2007). Pharmacological interaction between 3,4-methylenedioxymethamphetamine (ecstasy) and paroxetine: Pharmacological effects and pharmacokinetics. *The Journal of Pharmacology and Experimental Therapeutics, 323*, 954–962.

Feng, P., Feng, T., Chen, Z., & Lei, X. (2014). Memory consolidation of fear conditioning: Bi-stable amygdala connectivity with dorsal anterior cingulate and medial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 9*, 1730–1737.

Feng, P., Zheng, Y., & Feng, T. (2016). Resting-state functional connectivity between amygdala and the ventromedial prefrontal cortex following fear reminder predicts fear extinction. *Social Cognitive and Affective Neuroscience, 11*, 991–1001.

Fernandez Espejo, E. (2003). Prefrontocortical dopamine loss in rats delays long-term extinction of contextual conditioned fear, and reduces social interaction without affecting short-term social interaction memory. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 28*, 490–498.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 16311–16316.

Freese, J. L., & Amaral, D. G. (2009). Neuroanatomy of the primate amygdala. In P. J. Whalen, & E. A. Phelps (Eds.). *The human amygdala.* New York: The Guildford Press.

Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., ... Harrison, B. J. (2018). Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. Retrieved from *Neuroscience & Biobehavioral Reviews.* <https://www.sciencedirect.com/science/article/pii/S0149763417309600>.

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry, 21*, 500–508.

Galatzer-Levy, I. R., Bonanno, G. A., Bush, D. E. A., & Ledoux, J. E. (2013). Heterogeneity in threat extinction learning: Substantive and methodological considerations for identifying individual difference in response to stress. *Frontiers in Behavioral Neuroscience, 7*, 55.

Garfinkel, S. N., Abelson, J. L., King, A. P., Sripada, R. K., Wang, X., Gaines, L. M., & Liberzon, I. (2014). Impaired contextual modulation of memories in PTSD: An fMRI and psychophysiological study of extinction retention and fear renewal. *Journal of Neuroscience, 34*, 13435–13443.

Giovannini, M. G., Rakovska, A., Benton, R. S., Pazzagli, M., Bianchi, L., & Pepeu, G. (2001). Effects of novelty and habituation on acetylcholine, GABA, and glutamate release from the frontal cortex and hippocampus of freely moving rats. *Neuroscience, 106*, 43–53.

Gold, A. L., Shechner, T., Farber, M. J., Spiro, C. N., Leibenluft, E., Pine, D. S., & Britton, J. C. (2016). Amygdala-cortical connectivity: Associations with anxiety, development, and threat. *Depression and Anxiety, 33*, 917–926.

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

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron, 84*, 486–496.

Haaker, J., Gaburro, S., Sah, A., Gartmann, N., Lonsdorf, T. B., Meier, K., ... Kalisch, R. (2013). Single dose of l-dopa makes extinction memories context-independent and prevents the return of fear. *Proceedings of the National Academy of Sciences of the United States of America, 110*, E2428–E2436.

Haaker, J., Lonsdorf, T. B., & Kalisch, R. (2015). Effects of post-extinction l-DOPA administration on the spontaneous recovery and reinstatement of fear in a human fMRI study. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology, 25*, 1544–1555.

Hajós, M., Fleishaker, J. C., Filipiak-Reisner, J. K., Brown, M. T., & Wong, E. H. F. (2004). The selective norepinephrine reuptake inhibitor antidepressant reboxetine: Pharmacological and clinical profile. *CNS Drug Reviews, 10*, 23–44.

Hare, T. A., Tottenham, N., Galvan, A., Voss, H. U., Glover, G. H., & Casey, B. J. (2008). Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biological Psychiatry, 63*, 927.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. *Neuroreport, 11*,

43–48.

Hariri, A. R., Mattay, V. S., Tessitore, A., Fera, F., & Weinberger, D. R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. *Biological Psychiatry, 53*, 494–501.

Harrison, B. J., Fullana, M. A., Via, E., Soriano-Mas, C., Vervliet, B., Martínez-Zalacaín, I., ... Cardoner, N. (2017). Human ventromedial prefrontal cortex and the positive affective processing of safety signals. *NeuroImage, 152*, 12–18.

Hartley, C. A., & Phelps, E. A. (2010). Changing fear: The neurocircuitry of emotion regulation. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 35*, 136–146.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.), 293*, 2425–2430.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8*, 686–691.

Hermann, A., Keck, T., & Stark, R. (2014). Dispositional cognitive reappraisal modulates the neural correlates of fear acquisition and extinction. *Neurobiology of Learning and Memory, 113*, 115–124.

Hermann, A., Küpper, Y., Schmitz, A., Walter, B., Vaitl, D., Hennig, J., ... Tabbert, K. (2012). Functional gene polymorphisms in the serotonin system and traumatic life events modulate the neural basis of fear acquisition and extinction. *PLOS ONE, 7*, e44352.

Hermann, A., Stark, R., Blecker, C. R., Milad, M. R., & Merz, C. J. (2017). Brain structural connectivity and context-dependent extinction memory. *Hippocampus, 1*–7.

Hermann, A., Stark, R., Milad, M. R., & Merz, C. J. (2016). Renewal of conditioned fear in a novel context is associated with hippocampal activation and connectivity. *Social Cognitive and Affective Neuroscience, 11*, 1411–1421.

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Müller, C., & Lüthi, A. (2008). Switching on and off fear by distinct neuronal circuits. *Nature, 454*, 600–606.

Hikind, N., & Maroun, M. (2008). Microinfusion of the D1 receptor antagonist, SCH23390 into the IL but not the BLA impairs consolidation of extinction of auditory fear conditioning. *Neurobiology of Learning and Memory, 90*, 217–222.

Holtzman-Assif, O., Laurent, V., & Westbrook, R. F. (2010). Blockade of dopamine activity in the nucleus accumbens impairs learning extinction of conditioned fear. *Learning & Memory, 17*, 71–75.

Huff, N., Alba Hernandez, J., Fecteau, M., Zielinski, D., Brady, R., & LaBar, K. S. (2011). Revealing context-specific conditioned fear memories with full immersion virtual reality. Retrieved from *Frontiers in Behavioral Neuroscience, 5*.

Huh, K. H., Guzman, Y. F., Tronson, N. C., Guedea, A. L., Gao, C., & Radulovic, J. (2009). Hippocampal Erk mechanisms linking prediction error to fear extinction: Roles of shock expectancy and contextual aversive valence. *Learning & Memory, 16*, 273–278.

Izquierdo, A., Wellman, C. L., & Holmes, A. (2006). Brief uncontrollable stress causes dendritic retraction in infralimbic cortex and resistance to fear extinction in mice. *The Journal of Neuroscience, 26*, 5733–5738.

Jensen, J., McIntosh, A. R., Crawley, A. P., Mikulis, D. J., Remington, G., & Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron, 40*, 1251–1257.

Jin, J., & Maren, S. (2015). Fear renewal preferentially activates ventral hippocampal neurons projecting to both amygdala and prefrontal cortex in rats. *Scientific Reports, 5*, 8388.

Josselyn, S. A., Köhler, S., & Frankland, P. W. (2015). Finding the engram. *Nature Reviews Neuroscience, 16*, 521–534.

Jovanovic, T., & Norrholm, S. D. (2011). Neural mechanisms of impaired fear inhibition in posttraumatic stress disorder. Retrieved from *Frontiers in Behavioral Neuroscience, 5*.

Kafkas, A., & Montaldi, D. (2018). How do memory systems detect and respond to novelty? Retrieved from *Neuroscience Letters*.

Kahn, I., & Shohamy, D. (2013). Intrinsic connectivity between the hippocampus, nucleus accumbens, and ventral tegmental area in humans. *Hippocampus, 23*, 187–192.

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*, 9503–9511.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*, 679–685.

Kim, J. J., & Jung, M. W. (2006). Neural circuits and mechanisms involved in Pavlovian fear conditioning: A critical review. *Neuroscience and Biobehavioral Reviews, 30*, 188–202.

Kim, M. J., Loucks, R. A., Palmer, A. L., Brown, A. C., Solomon, K. M., Marchante, A. N., & Whalen, P. J. (2011). The structural and functional connectivity of the amygdala: From normal emotion to pathological anxiety. *Behavioural Brain Research, 223*, 403–410.

Kim, M. J., & Whalen, P. J. (2009). The structural integrity of an amygdala-prefrontal pathway predicts trait anxiety. *Journal of Neuroscience, 29*, 11614–11618.

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

Kinner, V. L., Wolf, O. T., & Merz, C. J. (2018). Cortisol increases the return of fear by strengthening amygdala signaling in men. Retrieved from *Psychoneuroendocrinology*.

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*, 317–325.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. Retrieved from *Frontiers in Systems Neuroscience, 2*.

Kroes, M. C. W., Tona, K.-D., den Ouden, H. E. M., Vogel, S., van Wingen, G. A., & Fernández, G. (2016). How administration of the beta-blocker propranolol before extinction can prevent the return of fear. *Neuropsychopharmacology, 41*, 1569–1578.

Kumaran, D., & Maguire, E. A. (2006). An unexpected sequence of events: Mismatch detection in the human hippocampus. *PLOS Biology, 4*, e424.

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*, 937–945.

Larrauri, J. A., & Schmajuk, N. A. (2008). Attentional, associative, and configural mechanisms in extinction. *Psychological Review, 115*, 640–676.

Lebron, K., Milad, M. R., & Quirk, G. J. (2004). Delayed recall of fear extinction in rats with lesions of ventral medial prefrontal cortex. *Learning & Memory, 11*, 544–548.

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

LeDoux, J. E. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology, 23*, 727–738.

LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 2871–2878.

Lee, J. L. C., Milton, A. L., & Everitt, B. J. (2006). Reconsolidation and extinction of conditioned fear: Inhibition and potentiation. *Journal of Neuroscience, 26*, 10051–10056.

Li, W., Howard, J. D., Parrish, T. B., & Gottfried, J. A. (2008). Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. *Science (New York, N.Y.), 319*, 1842–1845.

Li, S. S. Y., & McNally, G. P. (2014). The conditions that promote fear learning: Prediction error and Pavlovian fear conditioning. *Neurobiology of Learning and Memory, 108*, 14–21.

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*, 421–428.

Liechti, M. E., & Vollenweider, F. X. (2001). Which neuroreceptors mediate the subjective effects of MDMA in humans? A summary of mechanistic studies. *Human Psychopharmacology, 16*, 589–598.

Lin, H.-C., Mao, S.-C., Su, C.-L., & Gean, P.-W. (2009). The role of prefrontal cortex CB1 receptors in the modulation of fear memory. *Cerebral Cortex, 19*, 165–175.

Lindner, K., Neubert, J., Pfannmöller, J., Lotze, M., Hamm, A. O., & Wendt, J. (2015). Fear-potentiated startle processing in humans: Parallel fMRI and orbicularis EMG assessment during cue conditioning and extinction. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 98*, 535–545.

Linnman, C., Zeidan, M. A., Pitman, R. K., & Milad, M. R. (2012). Resting cerebral metabolism correlates with skin conductance and functional brain activation during fear conditioning. *Biological Psychology, 89*, 450–459.

Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron, 46*, 703–713.

Lisman, J. E., & Otmakhova, N. A. (2001). Storage, recall, and novelty detection of sequences by the hippocampus: Elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus, 11*, 551–568.

Lissek, S., Glaubitz, B., Uengoer, M., & Tegenthoff, M. (2013). Hippocampal activation during extinction learning predicts occurrence of the renewal effect in extinction recall. *NeuroImage, 81*, 131–143.

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *Journal of Neuroscience, 36*, 12677–12687.

Lonsdorf, T. B., Haaker, J., Fadai, T., & Kalisch, R. (2014). No evidence for enhanced extinction memory consolidation through noradrenergic reuptake inhibition—delayed memory test and reinstatement in human fMRI. *Psychopharmacology (Berl), 231*, 1949–1962.

Lonsdorf, T. B., Haaker, J., & Kalisch, R. (2014). Long-term expression of human contextual fear and extinction memories involves amygdala, hippocampus and ventromedial prefrontal cortex: A reinstatement study in two independent samples. *Social Cognitive and Affective Neuroscience, 9*, 1973–1983.

Mackie, K. (2005). Distribution of cannabinoid receptors in the central and peripheral nervous system. *Handbook of Experimental Pharmacology, 168*, 299–325.

Marek, R., Jin, J., Goode, T. D., Giustino, T. F., Wang, Q., Acca, G. M., ... Sah, P. (2018). Hippocampus-driven feed-forward inhibition of the prefrontal cortex mediates relapse of extinguished fear. *Nature Neuroscience, 21*, 384–392.

Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: Implications for fear conditioning, extinction and psychopathology. *Nature reviews. Neuroscience, 14*, 417–428.

Marin, M.-F., Song, H., VanElzakker, M. B., Staples-Bradley, L. K., Linnman, C., Pace-Schott, E. F., ... Milad, M. R. (2016). Association of resting metabolism in the fear neural network with extinction recall activations and clinical measures in trauma-exposed individuals. *The American Journal of Psychiatry, 173*, 930–938.

Marin, M.-F., Zsido, R. G., Song, H., Lasko, N. B., Killgore, W. D. S., Rauch, S. L., ... Milad, M. R. (2017). Skin conductance responses and neural activations during fear conditioning and extinction recall across anxiety disorders. *JAMA Psychiatry, 74*, 622–631.

Matsuo, N., & Mayford, M. (2008). Spine-type specific recruitment of newly synthesized AMPA receptors with learning. *Science (New York, N.Y.), 319*, 1104–1107.

McLaughlin, N. C. R., Strong, D., Abrantes, A., Garnaat, S., Cerny, A., O'Connell, C., ... Greenberg, B. D. (2015). Extinction retention and fear renewal in a lifetime obsessive–compulsive disorder sample. *Behavioural Brain Research, 280*, 72–77.

Mechias, M.-L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. *NeuroImage, 49*, 1760–1768.

Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: Awareness and response. *Brain Structure & Function, 214*, 535–549.

Menezes, J., Alves, N., Borges, S., Roehrs, R., de Carvalho Myskiw, J., Furini, C. R. G., ... Mello-Carpes, P. B. (2015). Facilitation of fear extinction by novelty depends on

dopamine acting on D1-subtype dopamine receptors in hippocampus. *Proceedings of the National Academy of Sciences of the United States of America, 112*, E1652–E1658.

Merz, C. J., Hamacher-Dang, T. C., Stark, R., Wolf, O. T., & Hermann, A. (2018). Neural underpinnings of cortisol effects on fear extinction. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 43*, 384–392.

Merz, C. J., Hamacher-Dang, T. C., & Wolf, O. T. (2014). Exposure to stress attenuates fear retrieval in healthy men. *Psychoneuroendocrinology, 41*, 89–96.

Merz, C. J., Hermann, A., Stark, R., & Wolf, O. T. (2014). Cortisol modifies extinction learning of recently acquired fear in men. *Social Cognitive and Affective Neuroscience, 9*, 1426–1434.

Merz, C. J., Tabbert, K., Schweckendiek, J., Klucken, T., Vaitl, D., Stark, R., & Wolf, O. T. (2012). Neuronal correlates of extinction learning are modulated by sex hormones. *Social Cognitive and Affective Neuroscience, 7*, 819–830.

Milad, M. R., Furtak, S. C., Greenberg, J. L., Keshaviah, A., Im, J. J., Falkenstein, M. J., ... Wilhelm, S. (2013). Deficits in conditioned fear extinction in obsessive-compulsive disorder and neurobiological changes in the fear circuit. *JAMA Psychiatry, 70*, 608–618.

Milad, M. R., Pitman, R. K., Ellis, C. B., Gold, A. L., Shin, L. M., Lasko, N. B., ... Rauch, S. L. (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biological Psychiatry, 66*, 1075–1082.

Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational neuroscience: Ten years of progress. *Annual Review of Psychology, 63*, 129–151.

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*, 446–454.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.

Mithoefer, M. C., Wagner, M. T., Mithoefer, A. T., Jerome, L., & Doblin, R. (2011). The safety and efficacy of ± 3,4-methylenedioxymethamphetamine-assisted psychotherapy in subjects with chronic, treatment-resistant posttraumatic stress disorder: The first randomized controlled pilot study. *Journal of Psychopharmacology (Oxford, England), 25*, 439–452.

Mithoefer, M. C., Wagner, M. T., Mithoefer, A. T., Jerome, L., Martin, S. F., Yazar-Klosinski, B., ... Doblin, R. (2013). Durability of improvement in post-traumatic stress disorder symptoms and absence of harmful effects or drug dependency after 3,4-methylenedioxymethamphetamine-assisted psychotherapy: A prospective long-term follow-up study. *Journal of Psychopharmacology (Oxford, England), 27*, 28–39.

Monfil, T., Vázquez Roque, R. A., Camacho-Abrego, I., Tendilla-Beltran, H., Iannitti, T., Meneses-Morales, I., ... Morales-Medina, J. C. (2018). Hyper-response to novelty increases c-Fos expression in the hippocampus and prefrontal cortex in a rat model of schizophrenia. *Neurochemical Research, 43*, 441–448.

Morrow, B. A., Elsworth, J. D., Rasmusson, A. M., & Roth, R. H. (1999). The role of mesoprefrontal dopamine neurons in the acquisition and expression of conditioned fear in the rat. *Neuroscience, 92*, 553–564.

Moscarello, J. M., & Maren, S. (2018). Flexibility in the face of fear: Hippocampal–prefrontal regulation of fear and avoidance. *Current Opinion in Behavioral Sciences, 19*, 44–49.

Mueller, D., Bravo-Rivera, C., & Quirk, G. J. (2010). Infralimbic D2 receptors are necessary for fear extinction and extinction-related tone responses. *Biological Psychiatry, 68*, 1055–1060.

Mueller, D., & Cahill, S. P. (2010). Noradrenergic modulation of extinction learning and exposure therapy. *Behavioural Brain Research, 208*, 1–11.

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

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*, 424–430.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An FMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience, 14*, 1215–1229.

Orsini, C. A., Kim, J. H., Knapska, E., & Maren, S. (2011). Hippocampal and prefrontal projections to the basal amygdala mediate contextual regulation of fear after extinction. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31*, 17269–17277.

Orsini, C. A., & Maren, S. (2012). Neural and cellular mechanisms of fear and extinction memory formation. *Neuroscience and biobehavioral reviews, 36*, 1773–1802.

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

Peri, T., Ben-Shakhar, G., Orr, S. P., & Shalev, A. Y. (2000). Psychophysiologic assessment of aversive conditioning in posttraumatic stress disorder. *Biological Psychiatry, 47*, 512–519.

Peters, J., Dieppa-Perea, L. M., Melendez, L. M., & Quirk, G. J. (2010). Induction of fear extinction with hippocampal-infralimbic BDNF. *Science (New York, N.Y.), 328*, 1288–1290.

Pezawas, L., Meyer-Lindenberg, A., Drabant, E. M., Verchinski, B. A., Munoz, K. E., Kolachana, B. S., ... Weinberger, D. R. (2005). 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: A genetic susceptibility mechanism for depression. Retrieved from *Nature Neuroscience*.

Phan, K. L., Fitzgerald, D. A., Nathan, P. J., Moore, G. J., Uhde, T. W., & Tancer, M. E. (2005). Neural substrates for voluntary suppression of negative affect: A functional magnetic resonance imaging study. *Biological Psychiatry, 57*, 210–219.

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*, 897–905.

Plendl, W., & Wotjak, C. T. (2010). Dissociation of within- and between-session extinction of conditioned fear. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*, 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 (Cold Spring Harbor, N.Y.), 12*, 399–406.

Quirk, G. J., Likhtik, E., Pelletier, J. G., & Paré, D. (2003). Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 23*, 8800–8807.

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

Quirk, G. J., Russo, G. K., Barron, J. L., & Lebron, K. (2000). The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 20*, 6225–6231.

Raczka, K. A., Mechias, M.-L., Gartmann, N., Reif, A., Deckert, J., Pessiglione, M., & Kalisch, R. (2011). Empirical support for an involvement of the mesostriatal dopamine system in human fear extinction. *Translational Psychiatry, 1*, e12.

Radulovic, J., & Tronson, N. C. (2010). Molecular specificity of multiple hippocampal processes governing fear extinction. *Reviews in the Neurosciences, 21*, 1–17.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory, 64*–99.

Resick, P. A., Williams, L. F., Suvak, M. K., Monson, C. M., & Gradus, J. L. (2012). Long-term outcomes of cognitive–behavioral treatments for posttraumatic stress disorder among female rape survivors. *Journal of Consulting and Clinical Psychology, 80*, 201–210.

Rougemont-Bücking, A., Linnman, C., Zeffiro, T. A., Zeidan, M. A., Lebron-Milad, K., Rodriguez-Romaguera, J., ... Milad, M. R. (2011). Altered processing of contextual information during fear extinction in PTSD: An fMRI study. *CNS Neuroscience & Therapeutics, 17*, 227–236.

Roy, A. K., Shehzad, Z., Margulies, D. S., Kelly, A. M. C., Uddin, L. Q., Gotimer, K., ... Milham, M. P. (2009). Functional Connectivity of the Human Amygdala using Resting State fMRI. *NeuroImage, 45*, 614.

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

Schiller, D., & Delgado, M. R. (2010). Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends in Cognitive Sciences, 14*, 268–276.

Schultz, W. (2010). Dopamine signals for reward value and risk: Basic and recent data. *Behavioral and Brain Functions: BBF, 6*, 24.

Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: An update and integration. *Neuroscience & Biobehavioral Reviews, 36*, 1740–1749.

Sehlmeyer, C., Dannlowski, U., Schöning, S., Kugel, H., Pyka, M., Pfleiderer, B., ... Konrad, C. (2011). Neural correlates of trait anxiety in fear extinction. *Psychological Medicine, 41*, 789–798.

Sehlmeyer, C., Schöning, 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*, e5865.

Senn, V., Wolff, S. B. E., Herry, C., Grenier, F., Ehrlich, I., Gründemann, J., ... Lüthi, A. (2014). Long-range connectivity defines behavioral specificity of amygdala neurons. *Neuron, 81*, 428–437.

Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction error governs pharmacologically induced amnesia for learned fear. *Science, 339*, 830–833.

Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learning & Memory, 21*, 580–584.

Shumake, J., Jones, C., Auchter, A., & Monfils, M.-H. (2018). Data-driven criteria to assess fear remission and phenotypic variability of extinction in rats. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 373*.

Shvil, E., Sullivan, G. M., Schafer, S., Markowitz, J. C., Campeas, M., Wager, T. D., ... Neria, Y. (2014). Sex differences in extinction recall in posttraumatic stress disorder: A pilot fMRI study. *Neurobiology of Learning and Memory, 113*, 101–108.

Sierra-Mercado, D., Padilla-Coreano, N., & Quirk, G. J. (2011). Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology, 36*, 529–538.

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

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

Soliman, F., Glatt, C. E., Bath, K. G., Levita, L., Jones, R. M., Pattwell, S. S., ... Casey, B. J. (2010). A genetic variant BDNF polymorphism alters extinction learning in both mouse and human. *Science, 327*, 863–866.

Spoormaker, V. I., Andrade, K. C., Schröter, M. S., Sturm, A., Goya-Maldonado, R., Sämann, P. G., & Czisch, M. (2011). The neural correlates of negative prediction error signaling in human fear conditioning. *NeuroImage, 54*, 2250–2256.

Spoormaker, V. I., Schröter, M. S., Andrade, K. C., Dresler, M., Kiem, S. A., Goya-Maldonado, R., ... Czisch, M. (2012). Effects of rapid eye movement sleep deprivation on fear extinction recall and prediction error signaling. *Human Brain Mapping, 33*, 2362–2376.

Sripada, R. K., King, A. P., Garfinkel, S. N., Wang, X., Sripada, C. S., Welsh, R. C., & Liberzon, I. (2012). Altered resting-state amygdala functional connectivity in men with posttraumatic stress disorder. *Journal of Psychiatry & Neuroscience: JPN, 37*, 241–249.

Tabbert, K., Merz, C. J., Klucken, T., Schweckendiek, J., Vaitl, D., Wolf, O. T., & Stark, R. (2010). Cortisol enhances neural differentiation during fear acquisition and extinction in contingency aware young women. *Neurobiology of Learning and Memory, 94*, 392–401.

Tonegawa, S., Liu, X., Ramirez, S., & Redondo, R. (2015). Memory engram cells have come of age. *Neuron, 87*, 918–931.

Tronson, N. C., Schrick, C., Guzman, Y. F., Huh, K. H., Srivastava, D. P., Penzes, P., ... Radulovic, J. (2009). Segregated populations of hippocampal principal CA1 neurons mediating conditioning and extinction of contextual fear. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 29*, 3387–3394.

Urry, H. L., van Reekum, C. M., Johnstone, T., Kalin, N. H., Thurow, M. E., Schaefer, H. S., ... Davidson, R. J. (2006). Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *Journal of Neuroscience, 26*, 4415–4425.

Utz, A., Thürling, M., Ernst, T. M., Hermann, A., Stark, R., Wolf, O. T., ... Merz, C. J. (2015). Cerebellar vermis contributes to the extinction of conditioned fear. *Neuroscience Letters, 604*, 173–177.

van Oort, E. S. B., van Cappellen van Walsum, A. M., & Norris, D. G. (2014). An investigation into the functional and structural connectivity of the Default Mode Network. *NeuroImage, 90*, 381–389.

van Well, S., Visser, R. M., Scholte, H. S., & 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*, 499–512.

VanElzakker, M., Fevurly, R. D., Breindel, T., & Spencer, R. L. (2008). Environmental novelty is associated with a selective increase in Fos expression in the output elements of the hippocampal formation and the perirhinal cortex. *Learning & Memory, 15*, 899–908.

VanElzakker, M., Kathryn Dahlgren, M., Caroline Davis, F., Dubois, S., & Shin, L. M. (2014). From Pavlov to PTSD: The extinction of conditioned fear in rodents, humans, and anxiety disorders. *Neurobiology of Learning and Memory, 113*, 3–18.

Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S. L., & Quirk, G. J. (2006). Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learning & Memory (Cold Spring Harbor, N.Y.), 13*, 728–733.

Visser, R. M., de Haan, M. I. C., Beemsterboer, T., Haver, P., Kindt, M., & Scholte, H. S. (2016). Quantifying learning-dependent changes in the brain: Single-trial multivoxel pattern analysis requires slow event-related fMRI. *Psychophysiology, 53*, 1117–1127.

Visser, R. M., Haver, P., Zwitser, R. J., Scholte, H. S., & Kindt, M. (2016). First steps in using multi-voxel pattern analysis to disentangle neural processes underlying generalization of spider fear. *Frontiers in Human Neuroscience, 10*, 222.

Visser, R. M., Kunze, A. E., Westhoff, B., Scholte, H. S., & Kindt, M. (2015). Representational similarity analysis offers a preview of the noradrenergic modulation of long-term fear memory at the time of encoding. *Psychoneuroendocrinology, 55*, 8–20.

Visser, R. M., Lau-Zhu, A., Henson, R. N., & Holmes, E. A. (2018). Multiple memory systems, multiple time points: How science can inform treatment to control the expression of unwanted emotional memories. Retrieved from *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*.

Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. *Nature Neuroscience, 16*, 388–390.

Visser, R. M., Scholte, H. S., & Kindt, M. (2011). Associative learning increases trial-by-trial similarity of BOLD-MRI patterns. *Journal of Neuroscience, 31*, 12021–12028.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature, 412*, 43–48.

Wilson, R. I., & Nicoll, R. A. (2002). Endocannabinoid signaling in the brain. *Science, 296*, 678–682.

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H.-J., & Düzel, E. (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45*, 459–467.

Wolosin, S. M., Zeithamova, D., & Preston, A. R. (2012). Reward modulation of hippocampal subfield activation during successful associative encoding and retrieval. *Journal of Cognitive Neuroscience, 24*, 1532–1547.

Xu, C., Krabbe, S., Gründemann, J., Botta, P., Fadok, J. P., Osakada, F., ... Lüthi, A. (2016). Distinct hippocampal pathways mediate dissociable roles of context in memory retrieval. *Cell, 167*, 961–972.e16.

Zelikowsky, M., Hast, T. A., Bennett, R. Z., Merjanian, M., Nocera, N. A., Ponnusamy, R., & Fanselow, M. S. (2013). Cholinergic blockade frees fear extinction from its contextual dependency. *Biological Psychiatry, 73*, 345–352.