Emotional valence interactions in amygdala circuits

Inauguraldissertation

Zur Erlangung der Würde eines Doktors der Philosophie vorgelegt der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel

Von

Lynda Demmou

aus Paris, Frankreich

Basel 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Andreas Lüthi
(Fakultätsverantwortlicher und Dissertationsleiter)

Prof. Dr. Rainer W. Friedrich
(Korreferent)

Prof. Dr. Jörg Schibler
(Dekan)

Basel, den 23.06.2015

TABLE OF CONTENTS

ABBREVIATIONS

BLA	Basolateral amygdala
BA	Basal nucleus of the amygdala
CeA	Central amygdala
CR	Conditioned response
CS	Conditioned stimulus
CS^-	Non-reinforced conditioned stimulus
CS^+_{ap}	Conditioned stimulus associated with an appetitive outcome
CS^+_{av}	Conditioned stimulus associated with an aversive outcome
FC	Fear conditioning
FX	Fear extinction
HTR	Hedonic taste reactivity
IL	Infra-limbic division of the medial prefrontal cortex
LA	Lateral nucleus of the amygdala
mPFC	Medial prefrontal cortex
PL	Pre-limbic division of the medial prefrontal cortex
PN	Pyramidal neuron
UR	Unconditioned response
US	Unconditioned stimulus
US_{ap}	Appetitive unconditioned stimulus
US_{av}	Aversive unconditioned stimulus

ABSTRACT

Survival of organisms crucially depends on their ability to adapt their behavior to changes in environmental circumstances. This adaptation to changes in the emotional significance of environmental cues is acquired through two different types of learning: either through conditioning, when animals learn the predictive relationship between environmental cues and biologically relevant outcomes or, through subsequent extinction learning, when the cue is not predictive anymore of the outcome. The amygdala is crucially involved in the learning processes regarding these changes in valence and contingency between stimuli and biologically relevant outcomes. Here we study at the single neuron level the representation and interaction of conditioning and extinction of opposite valences. We show that the basal nucleus of the amygdala encompasses distinct neuronal subpopulations responsible for learning specific changes in stimulus-outcome contingencies in a valence-dependent manner. We first identify basal amygdala neurons specifically responsive to either aversive conditioned cues, the so-called fear neurons, or exclusively to aversive extinguished cues, the fear extinction neurons. Subsequently, the development of a purely Pavlovian appetitive conditioning allowed us to determine that conditioning and extinction are encoded in a very similar manner in the appetitive and aversive domains. We identify appetitive neurons which are cue-responsive after appetitive conditioning and appetitive extinction neurons only responding to appetitive extinguished cues. The identification of these discrete neuronal populations which activity correlates with high and low emotional states raises the question of how conditioning and extinction of opposite valences are represented relative to each other in basal amygdala circuits. We address this question by combining sequential appetitive and aversive learning with chronic single unit recordings. Conditioning and extinction of opposite valences are mostly encoded in a segregated manner: conditioning neurons of one valence overlap neither with conditioning nor with extinction neurons of the opposite valence. In contrast, extinction neurons of opposite valence partially overlap, suggesting that extinction learning recruits valence-free and valence-independent mechanisms. Although the valence-specific conditioning and extinction neurons appear to be

spatially segregated, opposite valences interact with each other in time. We show that prior appetitive experience delays fear extinction learning without affecting fear conditioning. These behavioral findings are corroborated at the neuronal level by the insensitivity of fear neurons to prior appetitive experience whereas the activity of fear extinction neurons is reduced by prior appetitive experience. This demonstrates that prior emotional experience influences subsequent associative learning both at the behavioral and at the neuronal level. Finally, comparison of the basal amygdala responsiveness to aversive and appetitive cues reveals a strong aversive bias of amygdala circuits. Extinction resistant neurons, which post-conditioning cue-responsiveness is maintained after extinction learning, are responsible for this aversive bias. Like the other neuronal populations identified in this study, extinction-resistant neurons of opposite valence are mostly segregated. This suggests that these neurons participate in the maintenance of valence-specific memory traces after extinction learning and thus that aversive memories are more resistant to changes in stimulus-outcome contingency. Supporting this hypothesis, we also find a strong asymmetry of extinction training between aversive and appetitive valence: aversive extinction requiring much longer training than appetitive extinction.

INTRODUCTION

Emotions

By the crucial influence they exert on animal's behavior, emotions are at the core of the survival of organisms and species. Emotions are specific sets of psychological, physiological and behavioral reactions emerging from the interaction between an organism and its environment. Functionally emotions can be described as the combination of interoceptive and exteroceptive signals triggered by biologically relevant events and the associated behavioral strategies allowing animals to adapt to the circumstances[1–3]. In order to ensure their survival, it is indeed of fundamental importance for organisms to be able to detect cues in the environment which are associated with harm or danger, and cues associated with food resources or reproduction. In addition to this first detection step of salient events, assigning a specific valence to environmental stimuli is also crucial as it allows organisms to select from their behavioral repertoire appropriate responses and thus ensure avoidance of aversive outcomes and approach towards appetitive ones. Furthermore, emotions do not only participate to the survival of individuals, they also play an important role in the perpetuation of species. Communication of emotions by vocalizations, facial expression or postures allows animals to signal to their peers the presence of resources or danger, to signal their distress and call for help, and finally signal availability for reproduction.

Learning and memory

Learning corresponds to the process by which new memories are formed. Following this first step of memory acquisition, long-term memory storage is ensured by a phase of memory consolidation allowing for subsequent memory retrieval. The neuronal correlates of learning and memory consist of a wide variety of synergistic mechanisms ranging from the molecular scale to the mesoscopic scale.

At the mesoscopic scale, the different phases of encoding, storage and retrieval of the memory are thought to rely, at least partially, on different brain structures. The famous case study of Henry Molaison, widely known as patient HM, had a major impact in the delineation of the dependence on different brain areas of memories formed in the past compared to recently acquired memories. In the 1950s, Henry Molaison underwent a bilateral resection of large parts of the medial temporal lobes (including the hippocampal formation and adjacent structures) as an attempt to cure him of his epilepsy. As a result of the lobectomy, Henry Molaison suffered from a severe anterograde

amnesia and a temporally graded retrograde amnesia: he was able to remember events which occurred long before the brain surgery but was not capable of forming new memories of events occurring after the medial temporal lobes resection[4,5]. Recent technical advances have allowed to narrow down to the neuronal level the study of engrams. Taking advantage of the newly developed tools in optogenetics, Liu and colleagues demonstrated that the light-induced activation of hippocampal neurons recruited by the memory formation could induce a recall of the memory on the next day[6]. This study highlights the fact that memory formation and retrieval of recently formed memory rely, at least partially, on the same subset of neurons.

As most brain areas do not generate neurons after birth (with exception of the olfactory bulb and the dentate gyrus), memory formation is thought to rely on the combination of structural and molecular modifications which induce changes in connectivity and activity of pre-existing neuronal pools[7–10]. Changes in circuit connectivity result of the formation of new synapses or pruning[11]. In addition to structural changes participating in the rewiring of neuronal networks, changes in synaptic transmission rely on many molecular modifications such as the expression of neurotransmitter receptors, their trafficking to the synaptic cleft, their internalization or intracellular signalization cascades leading to gene modulations and expression of new molecules regulating neuronal protein expression linked with neurotransmitter detection, signaling cascades and action potential emission[7,12,13].

Despite the similarity of the cellular and molecular mechanisms thought to underlie the memory formation and retrieval among different brain areas, memory in itself is not a unitary concept. Similarly to the study of the temporal dynamic of memory acquisition and storage, insights on the existence of distinct types of memories differing by their content has stemmed from loss of function studies consecutive to brain lesions. After surgery Henry Molaison was still able to learn new motor skills but was not capable of remembering having learned them. This specific impairment in autobiographical memories leaving untouched other learning skills highlights the dependence on different brain regions of implicit memory (memory of motor skills and actions, like driving a car) and explicit memory (memory of facts or knowledge, like remembering where the car has been parked)[14]. In addition, distinct brain regions are thought to be involved in emotional memories. This dissociation between the emotional content of memory and explicit memory was demonstrated in Human by comparing the effects of restricted lesions of either the hippocampus

or the amygdala and lesions of both structures. A patient with amygdala-restricted lesions fails to acquire emotional memories while not showing impairment in explicit memories. In contrast, a patient with hippocampal lesions shows the opposite effect, i.e. deficit in explicit memory while emotional memory remains intact. Finally, a patient with lesions of both structures shows impairments for both emotional and explicit memories[14,15].

Emotional associative learning

Emotional associative learning is a specific type of memory formation initially described by Pavlov. The serendipitous discovery of this form of learning in the early 20th century had a major influence in the field of emotion research, in learning and memory and in psychology. While investigating the regulation of digestive processes, Pavlov made a groundbreaking observation of the transfer of innate behavioral responses from food to food predictors[16]. In these experiments, dogs were exhibiting salivation to food delivery, but gradually displayed salivation responses to the bell which preceded the food. This was the first description of emotional associative learning, a particular form of memory formation consisting in the establishment of a predictive relationship between a biologically relevant event (the unconditioned stimulus: US, i.e. the food) and environmental cues (conditioned stimuli: CS, i.e. the sound of the bell). Emotional associative learning is said to be contextual if the biologically relevant outcome is associated with the diffuse context or classical if a discrete event predicts the occurrence of the outcome. The stronger the contingency in space and time between the neutral elements of the context and the emotionally relevant event, the better predictor the context or the cues become of the emotionally salient event. An important distinction is also to be made between classical conditioning, in which a CS predicts the delivery of a US, from instrumental conditioning, in which contingency are established between the US delivery and the actions of the animal.

From the theoretical point of view, it has been proposed that the discrepancy between what the animal expects and the actual outcome drives the learning and the associative memory formation. The computation of this discrepancy has been captured by models such as the Rescorla-Wagner model which posits that the learning rate is proportional to the difference between an outcome and the prediction of this outcome[17]. This is conceptualized by the following equation: $\Delta V=\alpha\beta(\lambda-\sum V)$ where ΔV is the amount of learning, α the salience of the CS, β the speed of learning for a

given US, λ the actual outcome and ∑V the expectation. This model captures several important features of associative learning. First, if the received outcome is fully predicted (λ=∑V) no learning occurs, indicating that for learning to actually take place novelty is an important factor. Second, the difference between what the animal expects and what is obtained can either be positive or negative, leading respectively to either excitatory or inhibitory learning. Learning about the contingency between a stimulus and the delivery of an outcome (λ>∑V) corresponds to excitatory learning, or conditioning, whereas learning between the occurrence of a CS in the absence of the outcome (λ<∑V) corresponds to inhibitory learning or extinction learning.

Amygdala

The amygdala is a brain structure located deeply in the temporal lobe. It was first described in 1819 by the physiologist Karl Friedrich Burdach who, due to its shape, named it after the Greek root for almond. More than a century after Burdach initial anatomical description, a major advance in the unveiling of the amygdala function was achieved by loss of function studies. While performing lesions studies in rhesus monkeys as part of their research on the effects of mescaline, Klüver and Bucy described in 1937 profound emotional changes as the symptoms of bilateral temporal lobectomy, including the amygdala complex. Among other symptoms, amygdala-lesioned monkeys were unable anymore to exhibit behavioral reactions such as fear or anger[18]. These results were later corroborated in humans by the study of the Urbach-Wiethe disease, an extremely rare genetic disorder which often leads to the calcification of the medial temporal lobes inducing a necrosis of the amygdala complex[19]. Similar to the "emotional blindness" initially observed in the Klüver-Bucy syndrome, patients suffering from Urbach-Wiethe syndrome show impairment in the recognition of emotionally relevant stimuli. More recently, studies reporting symptoms of bilateral amygdala lesions in humans confirmed the link between amygdala function and fear processing. In a recent case study, an amygdala-lesioned patient was exposed to fearful stimuli, such as live snakes and spiders, but contrary to non-lesioned subjects, did not exhibit any fear reactions or experience any feeling of fear as assessed by subjective reports[20]. Interestingly these lesions studies highlight the fact that amygdala is important for both expression and feelings of fear.

Complementing loss of function studies, gain of function studies have also been performed in human and confirmed the role of the amygdala in emotional processing. As part of the pre-surgical evaluation of drug-resistant epilepsy, patients were implanted with intracerebral electrodes in the amygdala. The direct electrical stimulation of the amygdala induced emotions such as fear, sadness, anxiety but also feeling of happiness. Similarly to lesion studies previously described, the amygdala was shown to be important for both emotional subjective experience and psychophysiological responses[21].

However, these studies, by their lack of spatial resolution, refer at the amygdala as a single structure and fail to capture the fact that the amygdala is neither a functional nor a structural unit. Instead, the amygdala encompasses several nuclei differing by their cytoarchitecture, immunohistochemistry, connectivity and thus function[22,23]. Two main complexes constitute the amygdala: the basolateral nucleus (BLA) and the central nucleus (CeA). The BLA can be further divided into the lateral nucleus (LA) and the basal nucleus (BA). The BLA is a non-layered cortical-like structure, composed of 80% of glutamatergic projection neurons (PN). The remaining 20% of neurons consist in aspiny GABAergic interneurons[24] exhibiting a large variety of neurite morphology[25,26] and constitute several subclasses defined by the combinatorial expression of neuropeptides and calcium-binding proteins[22,27,28]. By making dense axonal baskets around the soma and the axon initial segment of pyramidal neurons[29], BLA interneurons regulate the generation of action potentials generation of PN and thus tightly control their output[30]. Recent publication using optogenetic manipulations of specific populations of BLA interneurons has shown that the molecular identity of these neurons is an important factor for their function in regulating fear learning[31].

Located medially to the BLA, the central amygdala is a striatal-like structure composed of four distinct subnuclei: the central capsular (CEc), the central intermediate (CEi), the central lateral (CEl) and the central medial amygdala (CEm)[32]. Contrary to the BLA complex, the central amygdala is mainly composed of GABAergic neurons[33].

The intrinsic and extrinsic connectivity of the amygdala relates to its pivotal role in integrating multisensory information in order to give rise to different types of behavioral strategies according to the circumstances. Beyond the difference in their cytoarchitecture, the BLA and the CeA also show specific connectivity pattern related to their function. Sensory information from different

modalities converges onto the amygdala at the level of the LA[34–38]. Two pathways provide sensory inputs to the LA: the direct pathway, consisting in afferences originating from the thalamus traveling through the internal capsule and the indirect pathway conveying information from the thalamus to the cortex and then to the lateral nucleus of the amygdala via the external capsule[36,39]. These two pathways are thought to convey information of increasing complexity depending on the involvement of the cortex[40]. The BLA complex also receives afferences from the hippocampus, the rhinal cortices and the prefrontal cortex[41–43]. The inputs from the hippocampal formation are thought to convey contextual information[44,45] while prefrontal ones would be implicated in behavioral flexibility[46,47]. Importantly, connections between the amygdala and these two brain regions are reciprocal suggesting the existence of long-range synergistic interactions allowing for context-dependent flexibility of emotional experience.

The central amygdala receives inputs from the BLA (Figure 1) and sends projections to brainstem structures such as the hypothalamus and periaqueductal gray. Because of its downstream position in the amygdala circuitry and its projection to brain regions controlling autonomic and neuroendocrine responses it has long been thought to be the output station of the amygdala complex implicated in the orchestration of behavioral and physiological responses.

This serial model of amygdala information flow has however been challenged by the description of direct sensory inputs onto the CeA[42] and by the resistance of certain types of emotional responses in BLA-lesioned animals[49,50]. This data suggests that depending on the circumstances the information processing in amygdala circuits can either use the serial or the parallel route.

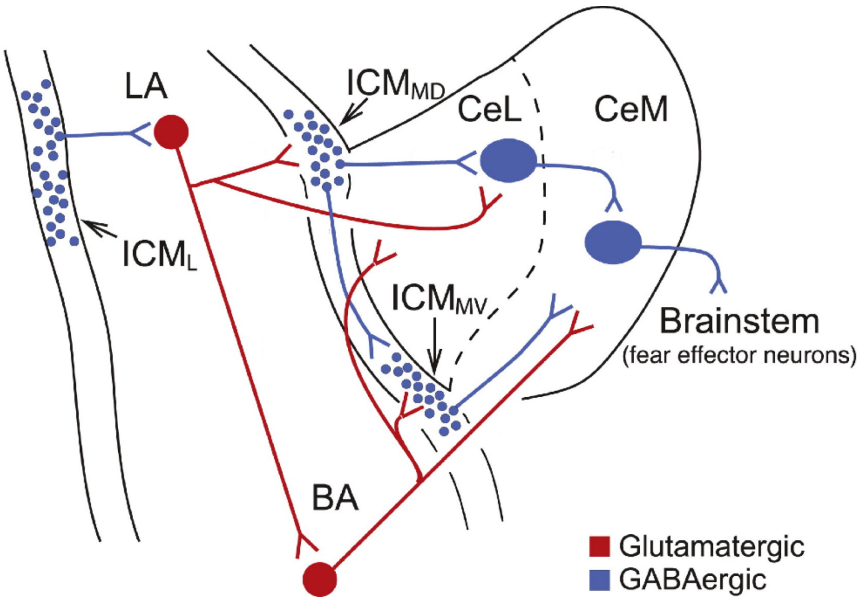

Figure 1. Amygdala intrinsic connectivity. Scheme of a coronal section of the rat amygdala representing the major internuclear connections (red: glutamatergic connection; blue GABAergic connection). LA: lateral nucleus; BA: basal nucleus; CeL: central lateral nucleus; CeM: central medial nucleus; ICM: intercalated cell masses, L: lateral, MD, mediodorsal, MV: medioventral. From Duvarci and Paré, 2014[48]

Fear conditioning

Fear conditioning (FC) is an associative learning process occurring when an otherwise neutral cue (the CS) is paired with an aversive outcome (the US) and thus gains an intrinsic aversive valence and/or leads to the expression of aversive conditioned responses (CR) when subsequently presented alone. A commonly used behavioral readout to assess fear conditioning in rodents is the freezing behavior which consists in the complete absence of movements of the animal, except for respiratory movements. Ethologically, this particular type of defensive behavioral response is admitted to prevent a pray from being detected by a close predator. However, this conditioned response differs drastically from the unconditioned responses (UR) exhibited at the time of the US delivery: upon footshock application (a commonly used US for FC in Rodents), rodents do not show freezing but flight responses. The selection of coping strategies thus highly depends on the emotional salience of a stimulus. Here, freezing emerges in response to the CS which predicts the footshock delivery while escape is observed upon the actual aversive outcome. At the end of this spectrum of defensive behavioral responses, animals can also exhibit fight responses. Thus the proximity from a predator or the imminence of an aversive outcome modulates the selection of appropriate behavioral responses.

For many decades, fear conditioning was the dominant model for studying the cellular and molecular underpinning of emotional associative learning in the amygdala. Several reasons participated in making fear conditioning such an influential model. First, historically amygdala activity was linked to fear expression[18,51,52]. Second, the robustness and the simplicity of the paradigm combined with the accessibility of behavioral measurements made it a model of choice for the laboratory. Finally, studying fear conditioning has a high translational potential as many psychopathological conditions (such as anxiety disorders and post-traumatic disorder) are related to disrupted fear regulations.

The critical role of the amygdala in fear processing has been initially demonstrated by permanent lesions studies[18,51] which showed that amygdala-lesioned monkeys are unable to express fear behavior. More recently, excitotoxic lesions (presenting the advantage of sparing fibers en passant) and reversible pharmacological inactivation (allowing for a better time resolution of the manipulation of brain activity and preventing compensatory effects) showed that the amygdala is not only necessary for fear expression but also for fear learning and memory. Using an olfactory

fear conditioning paradigm (in which an olfactory CS predicts a footshock), Cousens and Otto showed that BLA lesions performed prior to conditioning induced an impairment in conditioned freezing to the CS and to the context[53]. Additionally, lesions made after memory acquisition but before memory recall also induced impairment in conditioned freezing. This data suggests a crucial role of the BLA for both fear learning and fear memory retrieval. Considering the high temporal dynamic of memory formation processes it was, however, important to use more temporarily defined manipulation to disentangle the role of the amygdala in fear memory acquisition and consolidation. This was achieved by reversible inactivation of the BLA using the $GABA_A$ receptor agonist muscimol. Muscimol infusions directly before fear conditioning led to complete deficit in fear conditioning acquisition whereas post-training infusions had no effect on subsequent memory recall indicating that the activity in the BLA is required for the acquisition of fear conditioning but not for the consolidation of the memory[54].

In a similar way, gain of function studies have also participated in linking the amygdala function to fear expression. It was initially shown during the 1950s that amygdala electrical stimulation induces fear expression[52]. As suggested by anatomical studies, the LA is thought to be a site of convergence between CS and US sensory inputs. Taking advantage of the development of optogenetic approaches, Johansen and colleagues recently showed that indeed pairing a CS with light-induced activation of PN in the BLA is sufficient to produce conditioned fear responses[55].

Furthermore, numerous studies of the activity of the amygdala in humans and animal models have revealed a correlation between the amygdala activity and emotion expression, learning and memory.

Using fMRI (functional magnetic resonance imagery) Buchel and colleagues showed an increase in BOLD signal (blood-oxygen-level dependent, i.e. increased blood flow supporting a higher oxygenation of brain tissue thought to underlie increased energy demands upon brain activation) in the human amygdala during the presentation of cues previously associated with an aversive outcome[56]. Electrophysiological data also shows a correlation between BLA activity and fear memory. Local field potential recordings exhibit an enhancement of sensory-evoked responses in the BLA after fear conditioning[57]. Importantly, this FC-induced potentiation of sensory-evoked activity in the amygdala was shown to be specific of the CS associated with the US[58]. At the cellular level, several groups have used single unit recordings to show that individual neurons increase their

CS-responsiveness upon FC both in LA[59–62] and in the BA[63]. Finally, at the molecular level, FC induces N-methyl-aspartate receptor-dependent long-term potentiation in LA and infusion of NMDA antagonist in the LA impairs FC[64–66].

Appetitive conditioning

Appetitive conditioning is a learning process through which an organism forms a memory of the predictive relationships between its environment or its actions and rewarding outcomes. Despite the fact that emotional associative learning was initially described using classical appetitive conditioning[16], fear conditioning has dominated the field of classical conditioning while appetitive associative learning was mostly studied using instrumental paradigms.

However, the amygdala is not merely crucial for the processing of aversive experiences, it is also involved in the processing of positive emotions as demonstrated by the elicitation of a feeling of happiness by direct electrical stimulations of this brain structure[21]. In rodents, lesions of the BLA cause impairments in the association between the emotional valence of an outcome and its predictive cues, as assessed by second-order conditioning[67]. In these experiments, Rats were first trained to associated food delivery with a first CS. Then in a second stage of the experiment, the first CS was paired with a second CS but no reward was delivered anymore. Non-lesioned animals do exhibit appetitive conditioned responses to the second CS although it has never been paired with the food reward, suggesting that the first CS gained reinforcing power and emotional significance through its association with the US. In contrast, rats with BLA lesions fail to exhibit such transfer of appetitive responses from the first to the second CS indicating that the BLA is necessary for environmental cues to gain a positive emotional valence through their association with rewarding outcomes.

Similarly to fear conditioning, fMRI studies in humans have also shown increased BOLD signal in the amygdala in responses to appetitive CSs[68,69] confirming the involvement of the amygdala in emotional processing of both positive and negative valence. At the single cell level, Bermudez and Schultz demonstrated neurons in the monkey amygdala to not only be responsive for rewards but also to adapt their firing to the reward magnitude, the activity of some neurons increasing with the size of the reward while another neuronal population decreased its firing rate with increase of reward size[70]. This study also showed neurons responding to reward-predictive cues, a subset of the

US-responsive neurons increasing also their firing rate upon the presentation CSs associated with the reward delivery. This discrete BLA neuronal population, specifically responsive to reward-associated cues were first identified in 2006 by Paton and colleagues[71]. In this study, the activity in the BLA was monitored using single unit recordings in monkeys while they learned to associate one CS with the delivery of a liquid reward and another CS with the delivery of an aversive air puff. Importantly, neuronal responses were shown to be specific of the valence of the CSs as neurons were preferentially responding to appetitive cues (and not to aversively conditioned CSs) and encode the actual affective significance of the CS as their cue-evoked firing transfers to the other CS when the valence of the two cues is reversed. Additionally, these neuronal changes upon reversal of the valence of the two CSs precede the behavioral adaptation, suggesting a causal link between the activity of appetitive neurons and appetitive behavioral responses.

Other brain regions have been implicated in reward processing, such as the ventral tegmental area, the nucleus accumbens, the orbitofrontal cortex and the anterior cingulate[72]. Distributed circuitry among these different subregions is thought to subserve different aspects of reward processing (Figure 2)[73].

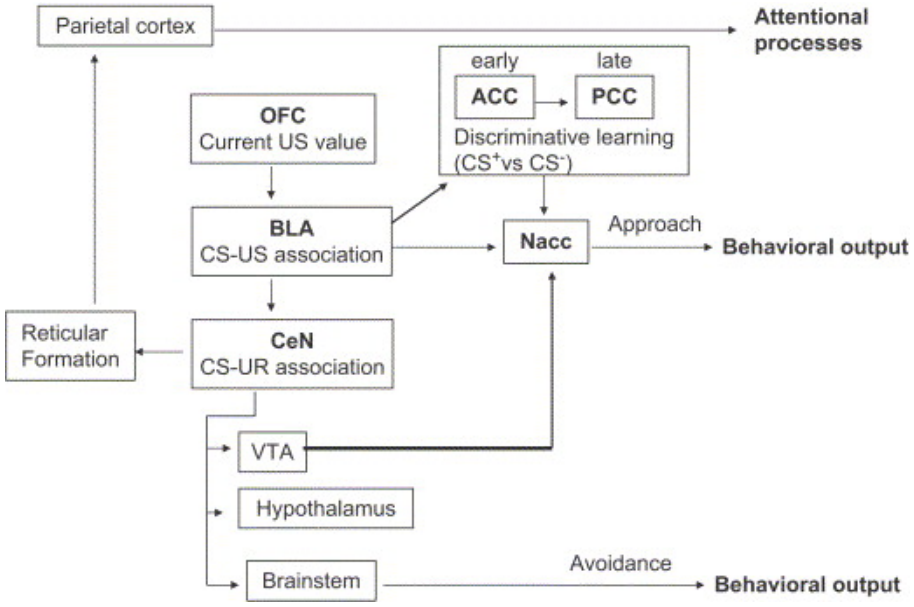

Figure 2. Model of the distributed brain network involved in appetitive. Findings summarized in this model emerge from both animals and human studies investigating functional connections in appetitive conditioning. OFC: orbito-frontal cortex; BLA: basolateral nucleus of the amygdala; CeN: central nucleus of the amygdala; VTA: ventral tegmental area; Nacc: nucleus accumbens; ACC: antero-cingulate cortex; PCC: posterior cingulate cortex. From Martin-Soelch, 2007[73]

In particular, the projections from the amygdala to the nucleus accumbens have been implicated in reward-seeking behavior, the amygdala being thought of signaling the relative valence of cues associated with rewarding USs to the nucleus accumbens which would act as a "limbic-motor interface" to produce approach behavior[73–75].

Fear extinction

Fear extinction is an inhibitory learning process leading to the gradual decline of fear conditioned responses as an adaptation to changes in the contingency between aversive events and the environment. Specifically, fear extinction occurs when a contextual or discrete cue previously associated with an aversive outcome is repeatedly presented in the absence of any adverse consequence.

Experimental psychology has made a major contribution to our understanding of fear extinction. Importantly, fear extinction does not lead to the destruction of the previously acquired fear memory trace but is rather a new associative learning between the environment and the absence of aversive outcome. This was demonstrated by the resurgence of fear responses with the passage of time after fear extinction (spontaneous recovery) or with a change of context (renewal) (Figure 3). Additionally, re-exposure to the US alone after completion of fear extinction leads to the reappearance of conditioned fear responses in the extinguished context (reinstatement) (Figure 3) and re-acquisition of fear conditioning consecutive to fear extinction takes place at a higher speed than the initial fear learning[76].

From an ethological point of view, it is indeed beneficial for organisms not to erase the fear memory trace but to rather form a new memory associating the previously aversive CS to learned safety in a specific context as the CS presented in a different context might still be predictive of an aversive outcome. However, this absence of contextual generalization of fear extinction has been a major challenge of psychotherapeutic approaches, such as exposure therapy, aiming at treating patients suffering from post-traumatic disorders. This behavioral evidence suggests that instead of erasing the previously acquired memory trace, fear extinction memory rather competes with fear memory in a context-dependent manner. However, it seems that depending on the circumstances, certain types of fear-conditioned responses can be completely abolished raising the possibility of

concomitant partial fear memory erasure and competition between fear and extinction memory traces (Figure 3)[77,78].

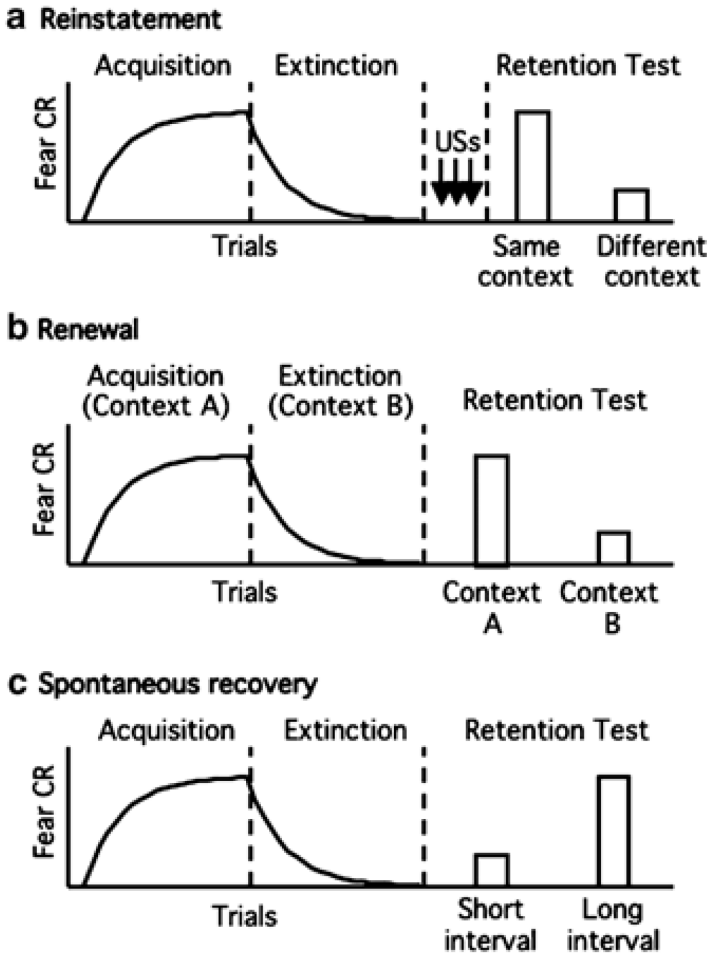

Figure 3. Extinguished fear responses recover under a variety of circumstances.
(**a**) *Reinstatement* occurs when unsignaled presentations of the US are interposed between the completion of extinction training and a subsequent retention test. Reinstatement is observed only if the USs are presented in the context in which the retention test will occur, indicating that the effect is context specific. (**b**) Extinction itself is context specific, as indicated by *renewal*. For example, if animals are fear conditioned in context A and extinguished in context B, they will exhibit extinction (i.e. little to no fear) if subsequently tested in context B, but they will show little evidence of extinction (i.e. renewed fear) if tested in context A. (**c**) *Spontaneous recovery* of extinguished fear responses occurs with the passage of time following extinction in the absence of any further training. The magnitude of recovery increases with the length of the extinction-to-test interval. From Myers and M Davis 2007[77]

Consistent with this two complementary mechanisms of fear extinction, changes in neuronal activity induced by FC are, at least partially, reversed by extinction learning while in parallel fear extinction also recruits specific neuronal circuits. At the single cell level, fear extinction causes a reduction in the cue-evoked firing rate of LA neurons which acquired CS-responsiveness through FC[59–62]. Importantly, neurons which lost their cue-responsiveness through fear extinction are still excited by fear-related cues, such as extinguished CSs presented in a context different from the extinction one (renewal). In addition, not all neuronal conditioned responses acquired through fear learning are reversed by fear extinction. Several groups have indeed observed in the amygdala fear extinction-resistant neurons, a specific class of neurons which maintains a high CS-responsiveness after fear extinction. This neuronal population is thought to contribute to the conservation of the fear memory trace in amygdala circuits after fear extinction[61–63].

In addition to the modulation of fear-induced neuronal responsiveness, fear extinction also recruits specific sets of neurons. Single unit recordings in the BA and in the LA identified neurons responding exclusively to fear extinguished cues[62,63]. Importantly, these fear extinction neurons are specifically responsive to extinguished cues as shown by the difference in their activity in a discriminative extinction paradigm where two cues were fear conditioned but only one was extinguished[63].

At the synaptic level, fear extinction relies on similar mechanisms than fear conditioning. BLA-injections of NMDA receptor antagonist indeed impairs fear extinction learning[79] whereas NMDA agonist injected in the BLA leads to facilitation of fear extinction[80].

In link with its high dependence on context, the fear extinction does not only rely on the activity of the amygdala but on the synergistic activity of a distributed network comprising the hippocampus and the infralimbic (IL) division of the medial prefrontal cortex (mPFC)[63,81]. Fear extinction neurons in the BA indeed receive inputs from the ventral hippocampus, thought to modulate the contextual dependency of fear extinction and project to the IL.

Appetitive extinction

The first description of appetitive extinction was made by Pavlov[16]. Similarly to fear extinction, it corresponds to the decline of appetitive conditioned responses due to change in the contingency between an appetitive outcome and previously associated cues. In Pavlov experiments, dogs conditioned to the sound of a bell for food delivery gradually decreased their behavioral responses (salivation) to the bell as it became less and less predictive of the food through repetitive presentations of the bell alone. As for fear extinction, behavioral studies have demonstrated that appetitive extinction does not lead to the erasure of the appetitive memory trace but rather corresponds to a context-dependent inhibitory learning leading to the coexistence of two competing memory traces[82].

Despite the early description of the phenomenon, the neuronal basis of appetitive extinction has been much less studied compared to the other forms of associative learning. Classical work from the 1950s has however demonstrated that permanent lesions of the amygdala complex in monkeys lead to impairment in appetitive extinction[51]. More recently, studies in rodents have confirmed the

involvement of the amygdala in appetitive extinction, excitotoxic lesions of the BLA in Rats leading to resistance to appetitive extinction training[83].

Only a few pharmacological studies have explored the role of the amygdala in appetitive extinction. Infusions of voltage-gated Na^+ channels blockers in the caudal division of the BA delays appetitive extinction learning in an instrumental task[84]. Extinction of cue-induced cocaine seeking behavior in Rats is impaired by post-training BLA injections of tetrodotoxin (voltage-gated Na^+ channel blocker), suggesting that the consolidation of appetitive extinction relies at least partially on BLA activity[85]. Because of its involvement in fear extinction process, Rhodes and Killcross tested the effect of excitotoxic lesions of the infralimbic division of the prefrontal cortex on appetitive extinction. They found appetitive extinction of instrumental CR (lever press associated with food delivery) to be insensitive to IL lesions but spontaneous recovery and reinstatement were increased in IL-lesioned rats compared to control animals[86]. Additionally, they showed in a subsequent publication that renewal (the resurgence of extinguished conditioned responses due to a contextual shift from the extinction context) was also higher in rats with IL lesions[86]. This data suggests, that as for fear extinction, IL may be implicated in the consolidation of appetitive extinction and in the flexibility of behavioral responses upon changes in environmental contingencies.

Measurements of the brain activity in relation to appetitive extinction have mostly focused on reward omission. Importantly, reward omission, and particularly unpredicted reward omission is very different from appetitive extinction as it does not rely on an active learning process. However, as described earlier, the discrepancy between expected and actual outcomes are thought to drive learning by operating as a teaching signal, thus reward omission-related neuronal activity could be seen as one of the first stages of the detection of changes in contingency between a predictive cue and a previously associated reward. The amygdala has been shown to be responsive to reward omission. In Humans, however, fMRI studies suggest that the amygdala although activated by reward omission, is more sensitive to reward delivery than to reward omission[87]. At the single neuron level, Belova and colleagues described a subset of neurons in the primate amygdala which exhibits similar excitatory responses to expected and non-expected rewards but was inhibited by reward omission[88].

Electrophysiological data on appetitive extinction *per se* is even scarcer than lesion and pharmacology studies. So far, only one study has identified neurons in the BLA which were

specifically responsive for port entries during the extinction period of an instrumental appetitive conditioning[89].

Emotional valence interactions

Evidence of interactions between learning episodes of opposite valences mostly stemmed from behavioral studies. During counterconditioning, a particular form of interference learning, animals are first trained to associate a specific cue with an outcome of one valence (a footshock for instance) and then to associate the same cue with an outcome of the opposite valence (like a food reward). Because the CS has been previously paired with an outcome of one particular valence it takes more time for animals to learn the association of this cue with the outcome of the opposite valence. At the behavioral level, it translates in a delay in the acquisition of valence-specific conditioned responses during the second conditioning episode[90–92]. Importantly, this effect is symmetrical for both valence reversals: prior appetitive learning delaying subsequent fear conditioning and *vice versa*[90]. This line of research relates to the so-called "opponent model" which postulates the existence of two distinct and mutually inhibiting neural systems underlying appetitive and aversive processing and which would be responsible for the behavioral expression of valence-specific responses according to environmental circumstances[93,94].

Consistent with the opponent model, segregated neuronal populations preferentially responding to either aversive or appetitive event have been identified in multiple brain areas. However, most studies investigating the neuronal representation of opposite valences also identified neurons responding similarly to both valences. Using TAI-FISH (a double-labeling technique based on the distinct time course of the mRNA and protein signals of the immediate early gene c-fos), Xiu et al. elegantly studied the segregation and convergence of appetitive (morphine) and aversive stimuli (foot-shock) in the limbic forebrain[95]. They found different patterns of interaction depending on the brain structures, some areas showing intermingled, some segregated and others overlapping representation of opposite valence (Figure 4).

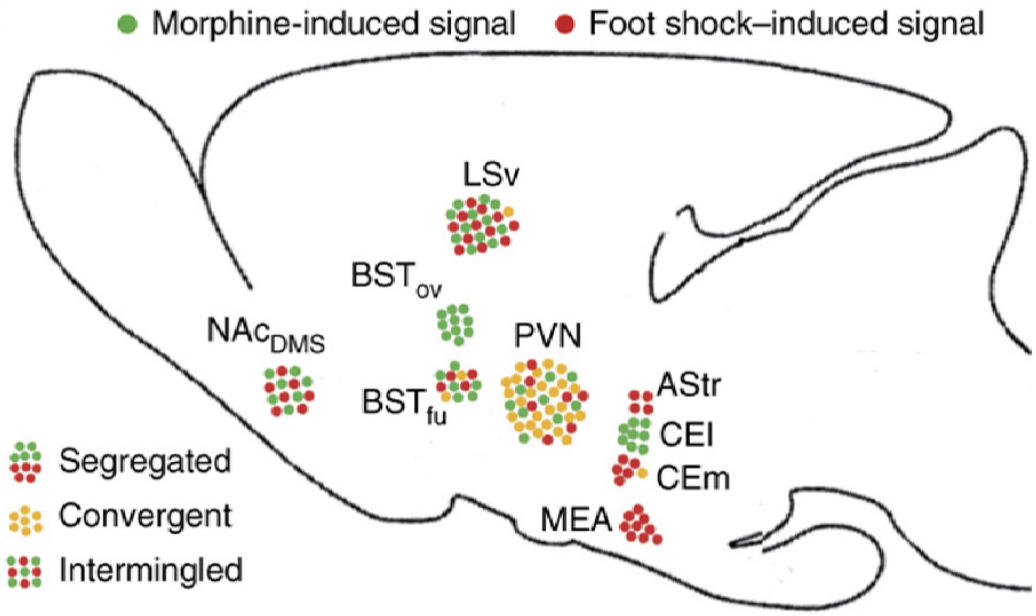

Figure 4. An emotional valence map in the forebrain. Summary of patterns of interaction between neural representations of morphine and foot shock in different regions of the limbic forebrain, as revealed by TAI-FISH (one dot represents 5 neurons counted from representative sections in each corresponding region). From Xiu, 2014[95]

Similarly, single unit recordings in the monkey amygdala did not only identified valence-specific neurons preferentially responding to either aversive or appetitive cue but also neurons responding similarly to both valences[71]. This two types of neuronal responses might reflect different components of the emotional experience. Indeed emotions can be described on a two-dimensional axis, one axis representing the valence and the second one representing the salience. The recruitment of common neuronal substrates by opposite valence could actually underlie valence-free mechanisms such as arousal or novelty detection whereas valence-specific neurons would participate in the computation of the specific emotional significance of biologically relevant events.

AIM OF THE THESIS

To adapt to circumstances and ensure their survival, animals need to attribute a relative emotional valence to environmental stimuli. This process relies on the interaction between the animal current state, its prior experiences and the external context. During my Ph.D., I studied the neuronal basis underlying the learning processes related to these changes in the emotional significance of environmental cues. Using conditioning and extinction of opposite valences, I investigated the encoding in amygdala circuits of changes in contingency and valence underlying behavioral adaptions. During the first part of my Ph.D., I focused on aversive learning and participated in a project identifying distinct BA neuronal populations contributing to fear conditioning or fear extinction. During the second part of my Ph.D., I studied Pavlovian appetitive conditioning and investigated the respective encoding of conditioning and extinction of opposite valences in amygdala circuits and the interaction between positive and negative emotional valences at the behavioral and neuronal levels.

SWITCHING ON AND OFF FEAR BY DISTINCT NEURONAL CIRCUITS

Switching on and off fear by distinct neuronal circuits[96]

Cyril Herry[1], Stephane Ciocchi[1], Verena Senn, Lynda Demmou, Christian Müller and Andreas Lüthi

[1]*These authors contributed equally to this work.*

Abstract

Switching between exploratory and defensive behavior is fundamental to survival of many animals, but how this transition is achieved by specific neuronal circuits is not known. Using the converse behavioral states of fear extinction and its context-dependent renewal as a model, we show that bi-directional transitions between states of high and low fear are triggered by a rapid switch in the balance of activity between two distinct populations of basal amygdala neurons. These two populations are integrated into discrete neuronal circuits differentially connected with the hippocampus and the medial prefrontal cortex. Targeted and reversible neuronal inactivation of the basal amygdala prevents behavioral changes without affecting memory or expression of behavior. Our findings indicate that switching between distinct behavioral states can be triggered by selective activation of specific neuronal circuits integrating sensory and contextual information. These observations provide a new framework for understanding context-dependent changes of fear behavior.

Introduction

The amygdala is a key brain structure mediating defensive behavior in states of fear and anxiety. Such states can be induced by classical auditory fear conditioning, in which an initially neutral auditory stimulus (the conditioned stimulus; CS) comes to elicit a fear response after pairing with an aversive foot-shock (the unconditioned stimulus; US). Subsequent repetitive presentations of the CS alone induce a progressive decrease in the fear response, a phenomenon called extinction. Whereas firing of amygdala neurons is critical for the retrieval of conditioned fear memories[97–102], their firing following the extinction of conditioned fear is thought to be constrained by local inhibitory circuits activated by the medial prefrontal cortex (mPFC)[102–107]. Converging evidence from animal studies indicates, however, that the basolateral complex of the amygdala (BLA), comprising the lateral (LA)

and the basal nucleus (BA), actively participates in fear extinction[77,79,108–111]. While fear extinction is an active learning process eventually leading to the formation of a consolidated extinction memory[77,111], it is a fragile behavioral state that is readily influenced by context[76,112]. Changing context results in the immediate recovery of the previously conditioned fear response, a process known as fear renewal[76,112]. *In vivo* pharmacological studies indicate that the hippocampus, which is reciprocally connected to the BLA[113], processes contextual information during fear conditioning, extinction, and renewal[44,45,76,114]. Thus, bi-directional changes in fear behavior during extinction and context-dependent renewal are likely to be encoded within a distributed network containing the BLA, the mPFC and the hippocampus, yet the neuronal circuits mediating such behavioral transitions are not known. In particular, this raises the question whether there are specialized circuits driving behavioral transitions in opposite directions.

To address this question, we used a combination of *in vivo* single unit recordings and targeted pharmacological inactivation in behaving mice. Because the BA is strongly connected to the hippocampus[113] and to the mPFC[41,115], and because extinction has previously been shown to induce the expression of the activity-dependent immediate early gene product c-Fos in BA neurons[116], we focused our study on this sub-nucleus. Here, we identify two distinct neuronal circuits differentially connected with the mPFC and the hippocampus, and show that a rapid switch in the balance of activity between those circuits specifically drives behavioral transitions without being necessary for memory storage or behavioral expression

Results

Distinct BA neurons encode fear and extinction

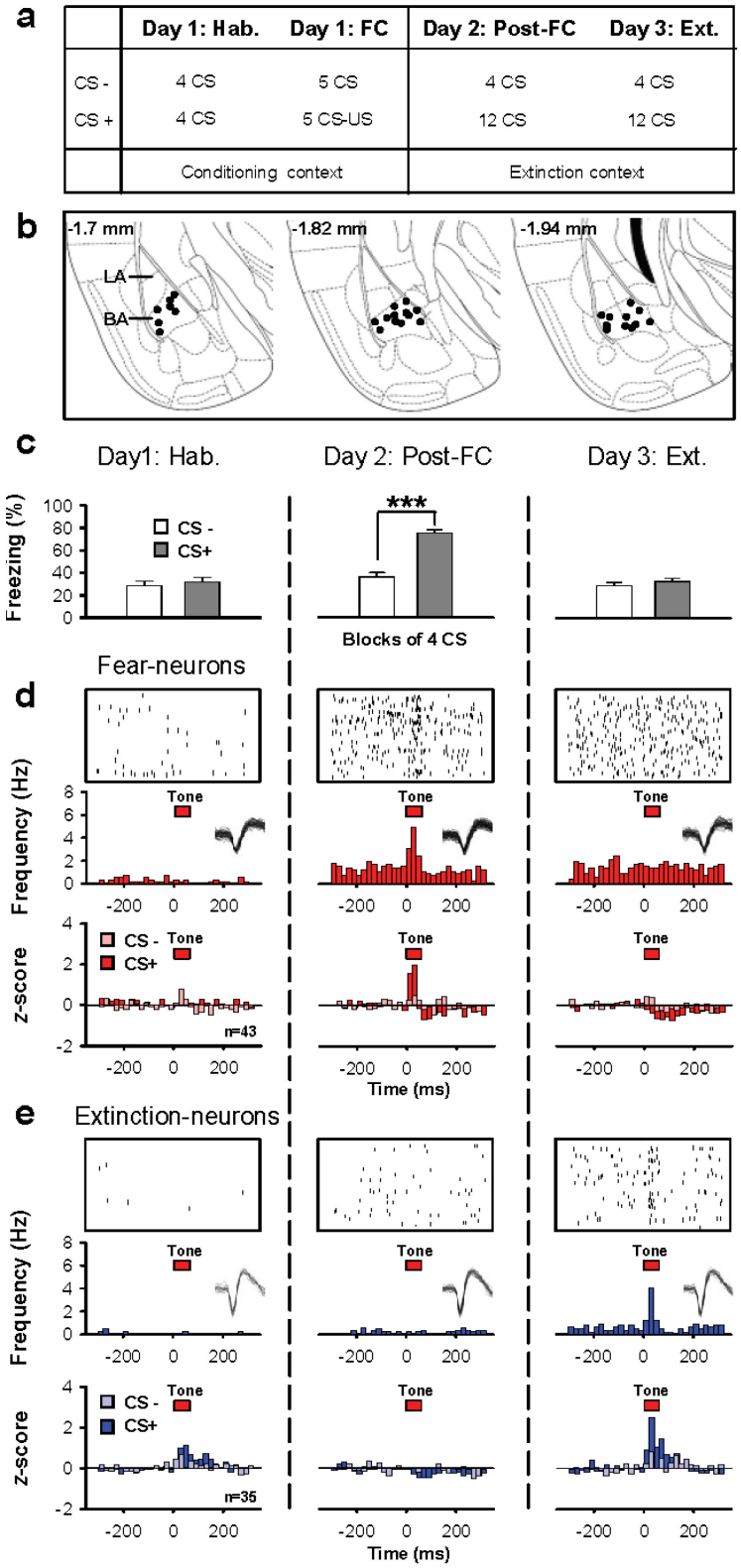

Figure 5. Distinct populations of BA neurons encode fear conditioning and extinction.
a, Experimental protocol. Hab.: habituation; FC: fear conditioning; Ext.: extinction. **b,** Coronal sections through the rostro-caudal extent of the amygdala showing the location of the recording sites in the BA. BA: basal nucleus of the amygdala; LA: lateral nucleus of the amygdala. **c,** Summary graph illustrating behavioral data. During habituation, mice (n = 30) exhibited equally low freezing levels in response to CS^+ and CS^- exposure. Twenty four hrs after fear conditioning, presentation of the CS^+ (CS 1 to 4), but not the CS^-, evoked significantly higher freezing levels. After extinction, both CS^+ (CS 9-12) and CS^- elicited low freezing levels. **d,e,** Raster plots (top) and peristimulus time histograms (middle) illustrating selective changes in CS^+-evoked firing of a representative fear- and extinction-neuron. Insets show superimposed spike waveforms recorded during habituation, after fear conditioning and after extinction. Bottom: Fear conditioning and extinction-induced changes in CS^+-evoked firing of fear- and extinction-neurons. Fear-neurons (n = 43 neurons from 22 mice) exhibited a selective increase in CS^+-evoked firing after fear conditioning ($P < 0.001$ vs. habituation or vs. CS^-), which was fully reversed upon extinction. In contrast, CS^+-evoked firing of extinction-neurons (n = 35 neurons from 20 mice) was selectively increased after extinction ($P < 0.001$ vs. post-FC or vs. CS^-). $^{***}P < 0.001$.

To examine plasticity of spike firing of individual BA neurons, C57Bl/6 mice were implanted with chronic recording electrodes and trained in a discriminative fear conditioning paradigm (Figure 5a).

During training mice learned to discriminate two auditory CSs of different frequencies. One CS (the CS^+) was paired with an aversive foot-shock (unconditioned stimulus; US), while the second CS (CS^-) was not paired. Twenty four hours after fear conditioning, mice (n = 30) exhibited a selective increase in fear behavior (as measured by freezing) when exposed to the CS^+ in a different context (Figure 5c). Extinction of conditioned fear behavior was induced by exposing mice to 24 CS^+ presentations in the absence of any aversive stimuli. After extinction training, CS^+-induced freezing behavior was reduced back to pre-conditioning levels, and did not differ from CS^--induced freezing (Figure 5c).

Analysis of changes in CS^+- and CS^--evoked spike firing during extinction training revealed that BA neurons (259 recorded units; Figure 5b) could be divided into distinct functional classes. Consistent with previous reports[117,118], we found a class of neurons ("fear-neurons"; n = 43 neurons, 22 mice; 17% of recorded units) exhibiting a selective increase in CS^+-evoked spike firing during and after fear conditioning (Figure 5d; Figure 6; Table 1).

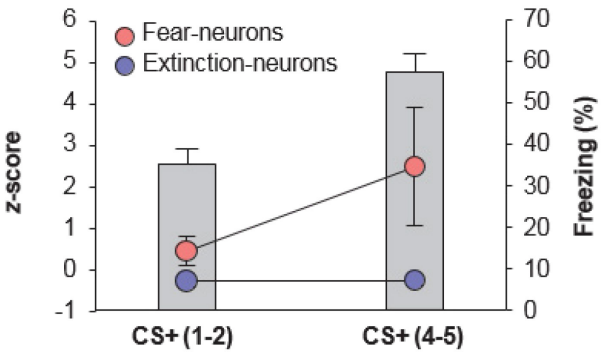

Figure 6. Changes in CS-evoked activity during fear conditioning.
Summary graph illustrating changes in freezing behavior (grey bars), and CS-evoked activity of fear-neurons (red circles) and extinction-neurons (blue circles). Comparing the first two CSs (CS 1-2) with the last two CSs (CS 4-5) reveals that increased freezing behavior (CS 1-2: 35 ± 4% of time; CS 4-5: 58 ± 4% of time) was associated with enhanced CS-evoked activity in fear neurons (n = 43 neurons from 22 mice, z-score, CS 1-2: 0.41 ± 0.35; CS 4-5: 2.45 ± 1.42), but not in extinction neurons (n = 35 neurons from 20 mice, z-score, CS 1-2: -0.31 ± 0.15; CS 4-5: -0.29 ± 0.11).

Subsequent extinction completely abolished this increase and converted it into a CS^+-evoked inhibition (Figure 5d). On average, spontaneous activity of fear neurons was not affected by fear conditioning or extinction (Table 1). Thus, fear conditioning-induced behavioral discrimination between the CS^+ and the CS^-, and its reversal by extinction, was accurately reflected at the neuronal level by the discriminative and reversible activity of fear-neurons.

	Habituation			Post-FC			Extinction			Percent of units (%)
	z-score		Spont. Freq. (Hz)	z-score		Spont. Freq. (Hz)	z-score		Spont. Freq. (Hz)	
	CS-	CS+		CS-	CS+		CS-	CS+		
Fear neurons	0.46±0.44	0.38±0.25	2.2±0.5	0.37±0.21	**1.78±0.28*****	1.8±0.3	0.42±0.21	-0.30±0.10	1.6±0.3	16.6
Extinction neurons	0.61±0.41	0.57±0.20	1.5±0.3	-0.04±0.13	-0.17±0.14	1.1±0.3	0.51±0.18	**1.71±0.35*****	1.5±0.3	13.5
Extinction-resistant neurons	1.49±0.37	0.84±0.19	2.3±0.5	1.29±0.24	**2.64±0.45****	**3.6±0.7*****	1.54±0.30	**2.49±0.44****	**4.2±0.8****	25.5

Table 1: Summary of units recorded in BA.
This table summarizes changes in CS-induced neuronal activity (z-scores) and in spontaneous activity across behavioral sessions. Post-fear conditioning (post-FC) values were obtained using the first 4 CS+ presentations on day 2. Post-extinction (extinction) values were obtained using the last 4 CS+ presentations on day 3. Spontaneous activity was measured during the 500 ms preceding CS stimulation. Statistical comparisons: z-scores, CS+ vs. CS– within each behavioral session; spontaneous activity, post-FC and extinction vs. habituation. *P < 0.05, **P < 0.01, ***P < 0.001.

During extinction training, another class of neurons emerged. In contrast to fear-neurons, "extinction-neurons" (n = 35 neurons, 20 mice; 14% of recorded units) did not show any increase in CS-evoked responses during or after fear conditioning, but rather a slight reduction (Figure 5e). However, subsequent extinction training induced a marked and selective increase in CS^{+}-evoked activity in these neurons (Figure 5e), without any changes in spontaneous activity. Plotting extinction-induced changes in z-score for individual fear- and extinction-neurons revealed that the two populations were separated in a bi-modal distribution (Figure 7).

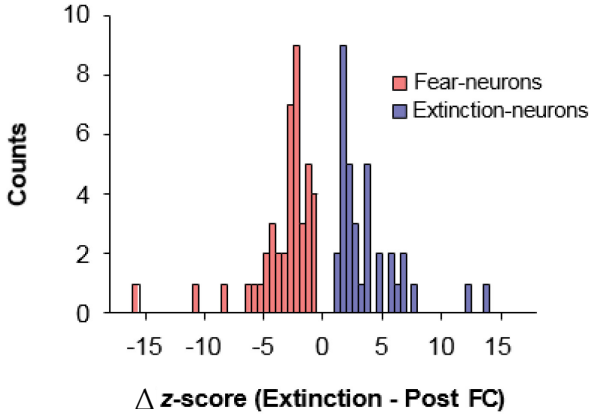

Figure 7. Extinction-induced changes in CS-evoked activity reveal a bimodal distribution of fear- and extinction-neurons.
Histogram representing the extinction-induced changes in the CS+-evoked neuronal activity ($z$score) of individual fear-neurons (n = 43) and extinction-neurons (n = 35). A negative Δz-score value indicates a preferential activation after fear conditioning, whereas a positive Δz-score value indicates a preferential activation after extinction. Fear- and extinction-neurons formed two well-separated populations.

The remaining neurons did not exhibit any changes in activity during extinction (Table 1). Thus, changes in CS^{+}-evoked firing of fear- and extinction-neurons were oppositely correlated with behavioral extinction.

While these results demonstrate a specific activation of fear- and extinction-neurons by a given CS, they do not address the question whether individual extinction-neurons can function as fear-neurons for another CS, or vice versa. We therefore trained mice in a discriminative extinction paradigm (Figure 8a).

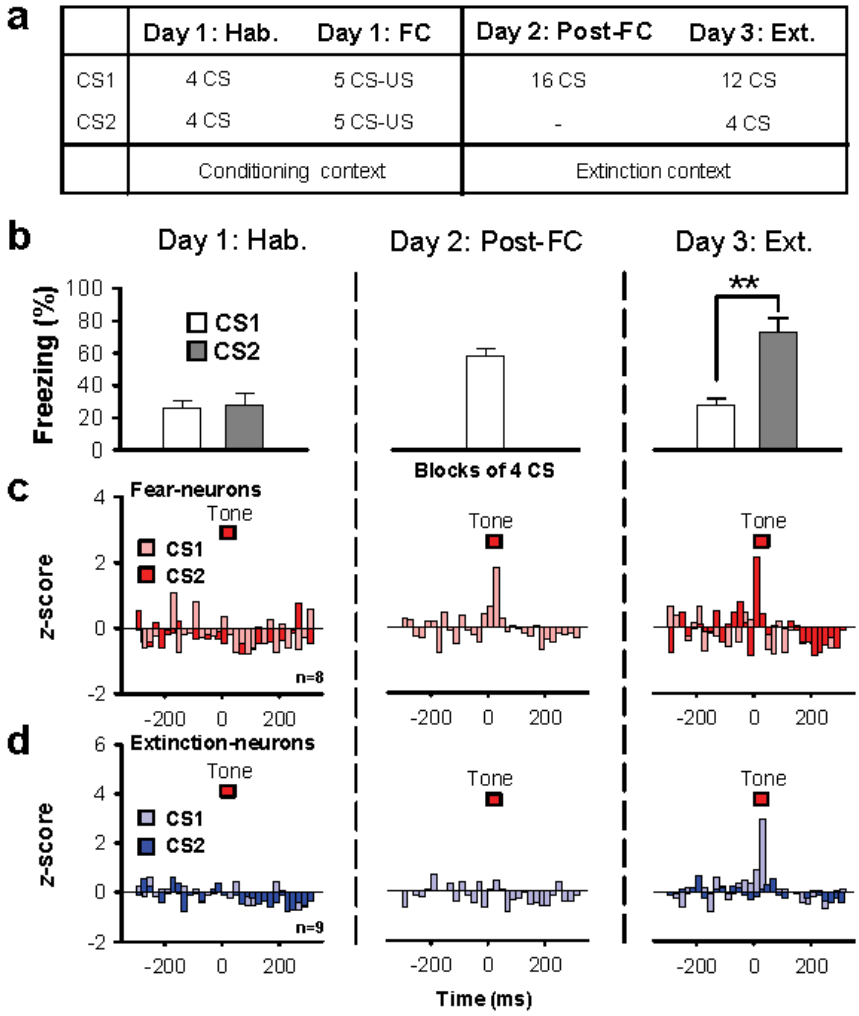

Figure 8. Fear- and extinction-neurons discriminate stimuli with different emotional significance.
a, Experimental design for discriminative extinction training. Initially, animals were fear conditioned to two distinct CSs (CS1 and CS2). Subsequently, only one CS (CS1) was extinguished. **b,** Summary of behavioral data. During habituation, mice (n = 6) exhibit equally low freezing levels in response to CS1 and CS2 exposure. After fear conditioning, presentation of the CS1 (CS 1 to 4) evokes significantly increased freezing levels. After extinction to CS1, CS1 exposure (CS 9 to 12) elicits low freezing levels, while CS2-evoked freezing behavior remains high. **c,** Fear conditioning- and extinction-induced changes in CS1- and CS2-evoked firing of fear-neurons (n = 8 neurons from 3 mice). Twenty four hrs after fear conditioning (day 2), fear-neurons exhibited increased firing in response to CS1 stimulation. After extinction of CS1, only CS2 stimulation elicited significant firing (day 3)($P < 0.05$ vs. CS1). **d,** Fear conditioning- and extinction-induced changes in CS1- and CS2-evoked firing of extinction-neurons (n = 9 neurons, 3 mice). After fear conditioning (day 2), extinction-neurons did not respond to CS1 stimulation. After extinction of CS1, only CS1 stimulation elicited significant firing (day 3) ($P < 0.05$ vs. CS2).

In this paradigm, two different CSs (CS1 and CS2) were first fear-conditioned, followed by extinction of only one of them (CS1). At the end of extinction, mice exhibited selective freezing

behavior when exposed to the non-extinguished CS2 (Figure 8b). Fear-neurons and extinction-neurons were identified during fear conditioning and extinction of CS1 according to the same criteria as described above, and CS1- and CS2-evoked spike firing was compared at the end of extinction. While individual extinction-neurons (n = 9 neurons, 3 mice) responded to the extinguished CS (CS1), but not to the non-extinguished CS (CS2), fear-neurons (n = 8, 3 mice) only fired following CS2 exposure, but remained unresponsive to the CS1 (Figure 8c, d). These observations confirm that individual fear-neurons and extinction-neurons represent functionally distinct classes of neurons that can discriminate between extinguished and non-extinguished stimuli.

In addition to the BA, we also recorded from 38 neurons in the LA which represents the main target of sensory afferents from thalamus and cortex[101]. In keeping with previous studies[61,97], we did not observe any LA neuron in which CS^+-evoked firing increased during extinction. Although we cannot exclude the existence of such neurons in LA, this may suggest that extinction-neurons are specific for the BA, where they represent 14% of all recorded neurons.

Activity balance predicts behavior

Comparing the averaged time courses of CS-evoked activity of fear- and extinction-neurons during the acquisition of behavioral extinction indicated that significant behavioral changes occurred after the activity scores of the two populations of neurons crossed over (Figure 9a). The largest changes in CS-evoked activity for both fear- and extinction-neurons occurred between the 3rd and the 4th block of extinction training, which are separated by 24 h, suggesting that an overnight consolidation process may be required. To further investigate the exact time point during extinction learning at which fear- and extinction-neurons displayed a significant change in activity we applied a change point analysis algorithm[119]. Change point analysis identifies the trial(s) exhibiting a significant change in neuronal activity or freezing behavior relative to the preceding trials. This analysis confirmed that changes in neuronal activity precede behavioral changes, and revealed that the activity of extinction-neurons started to increase one trial before the activity of fear-neurons began to decline (Figure 9b, c). Plotting activity changes of single fear- and extinction-neurons recorded in the same animal showed that the sequence of events is the same in an individual animal, and that such changes occur abruptly in an all-or-none manner (Figurer 9c).

This is consistent with the idea that behavioral changes are driven by sequential switches in the activity of two distinct neuronal circuits.

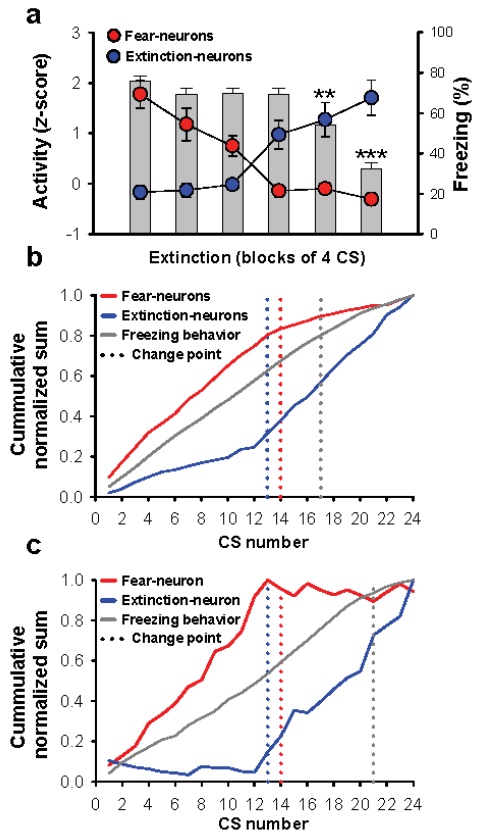

Figure 9. Sequential switches in neuronal activity precede behavioral changes. a, Averaged time courses of freezing behavior (grey bars; n = 30 mice) and neuronal activity (z-scores) of BA fear-neurons (red circles; n = 43) and extinction neurons (blue circles; n = 35) during extinction training. Significant behavioral changes (i.e. decreased freezing levels) occurred after activity scores of fear- and extinction neurons have crossed over. **b,** Change point analysis confirms that changes in neuronal activity preceded behavioral changes, and demonstrates that the activity of extinction neurons started to increase one trial before the activity of fear-neurons changed. Plot represents the cumulative sums of the averaged and normalized z-scores of fear- and extinction neurons, and freezing behavior during extinction training. Change points are indicated by dotted lines. **c,** Normalized cumulative sums of the z-scores of a single fear-neuron and a single extinction neuron recorded in the same animal together with the corresponding freezing behavior during extinction training. Change point analysis reveals that the extinction neuron abruptly switched on one trial before the fear neuron switched off. Changes in neuronal activity preceded behavioral changes. Change points are indicated by dotted lines.

Rapid reversal of activity during fear renewal

To test whether the activity of fear- and extinction-neurons represents the same behavioral values in a different paradigm, we analyzed renewal of extinguished fear behavior and associated changes in CS-evoked spike firing. In order to make sure that extinction memory was stably consolidated, mice (n = 15) were tested for extinction memory 7 days after extinction training in the same context in which extinction training occurred (Figure 10a).

After successful recall of extinction memory (Figure 10b), mice were transferred to the context in which they had been initially fear conditioned. Changing context resulted in a modest, but significant increase in baseline freezing levels due to contextual fear conditioning (Figure 11), and in a full renewal of the original cued fear memory (Figure 10b).

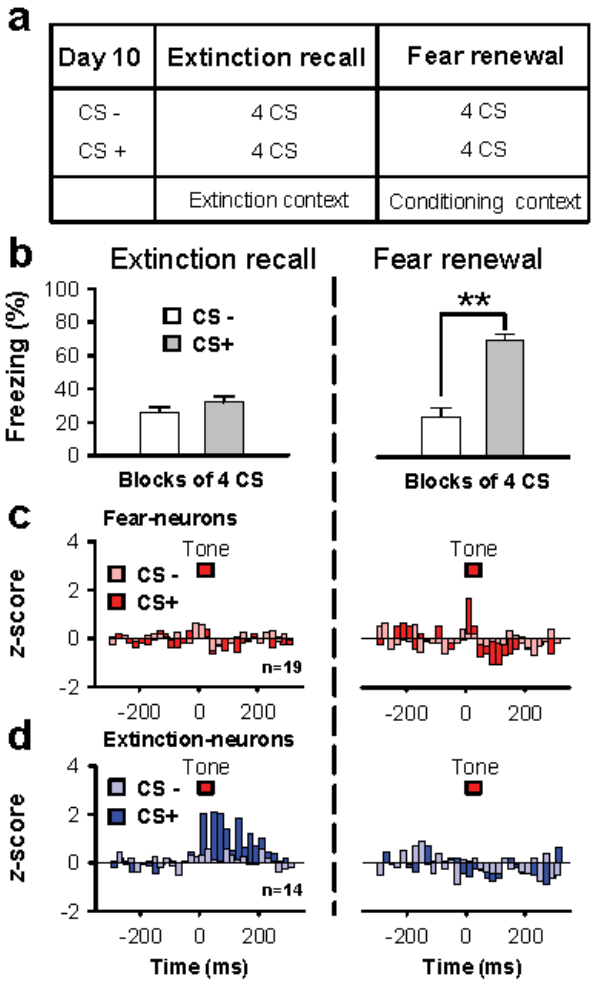

Figure 10. Context-dependent fear renewal induces rapid reversal of neuronal activity patterns.
a, Experimental protocol. **b,** Summary of behavioral data. Seven days after extinction, extinction memory was tested in the same context in which extinction training took place (n = 15 animals). Both CS^+ and CS^- elicit low freezing behavior. Subsequently, mice were placed back into the context in which fear conditioning took place. In this context, exposure to the CS^+ evoked significantly more freezing than CS^- stimulation. **c,** Context-dependent changes in CS^+-evoked firing of fear-neurons (n = 19 neurons from 9 mice). Fear-neurons exhibit a context-dependent increase in CS^+-evoked firing in the fear conditioning context where freezing levels are high ($P < 0.05$ vs. extinction context and vs. CS^-). **d,** Extinction-neurons (n = 14 neurons, 8 mice) show the opposite pattern. While CS^+-exposure elicits strong firing in the extinction context ($P < 0.05$ vs. fear conditioning context and vs. CS^-), extinction-neurons do not show any CS^+-evoked responses in the fear conditioning context. $^{**}P < 0.01$.

During recall of extinction memory in the extinction context, presentation of the CS^+ induced a selective activation of extinction-neurons (n = 14, 8 mice) with no effect on fear neurons (n = 19, 9 mice; Figure 10c, d). Thus, activation of extinction neurons by an extinguished CS is not a transient phenomenon, but remains stable for at least one week. After placing the animals in the fear conditioning context, increased CS^+-evoked freezing behavior was associated with a complete reversal of spiking activity at the cellular level. While extinction-neurons stopped responding to CS^+ stimulation, fear-neurons exhibited a significant and selective increase in CS^+-evoked spike firing (Figure 10d). Extinction-resistant neurons were not significantly activated during renewal (not shown). Thus, a switch in the balance of activity between fear- and extinction-neurons not only reflects extinction, but also parallels rapid context-dependent renewal of conditioned fear responses.

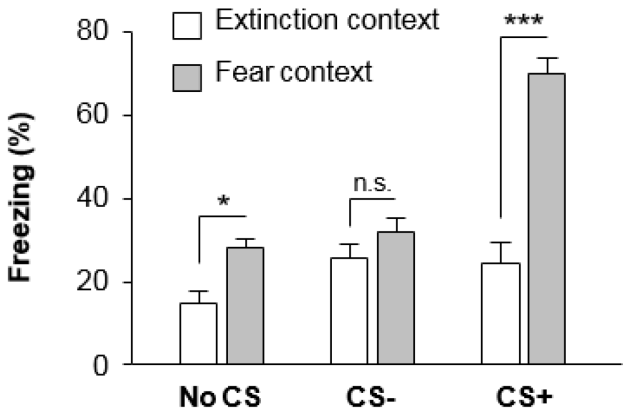

Figure 11: Context-dependent freezing during fear renewal.
Seven days after extinction, mice (n = 15 animals) were exposed to the CS+ and to the CS- in the extinction context and in the context in which fear conditioning took place. In the extinction context, both the CS+ and the CS– elicited low freezing behavior (CS–: 26 ± 3% of time; CS+: 32± 3%, P = 0.128 vs. CS–, P = 0.513 vs. extinction; same data as shown in figure 10). In the fear conditioning context, mice exhibited a modest, but significant increase in baseline freezing levels due to contextual fear conditioning (extinction context: 15 ± 3% of time; fear conditioning context: 28 ± 2%, P < 0.05), which was not significantly different from CS–-induced freezing. In this context, exposure to the CS+ evoked significantly more freezing than CS– stimulation (CS–: 24 ± 5% of time; CS+: 70 ± 4%, P < 0.01 vs. CS– , P < 0.001 vs. extinction recall; same data as shown in figure 10). *P < 0.05, ***P < 0.001.

Differential long-range connectivity

We next addressed the question whether fear-neurons and extinction-neurons were anatomically segregated. Comparing the location of electrolytic lesions made by the electrodes from which fear- and extinction-neurons were recorded did not provide any evidence for anatomical segregation (Figure 12a). As a complementary approach, we compared the anatomical distribution of BA neurons activated during exposure to an extinguished or to a non-extinguished CS using the immediate early gene product c-Fos as an activity-marker. Given the similar numbers of extinction and fear neurons, one would predict that an extinguished and a non-extinguished CS should induce c-Fos-expression in an equal number of BA neurons with an overlapping anatomical distribution. Consistent with this, we found no difference in the density and anatomical distribution of c-Fos-positive neurons in animals exposed to an extinguished and a non-extinguished CS (Figure 12c, d). Together, these results suggest that BA fear- and extinction-neurons are intermingled in a salt and pepper-like fashion.

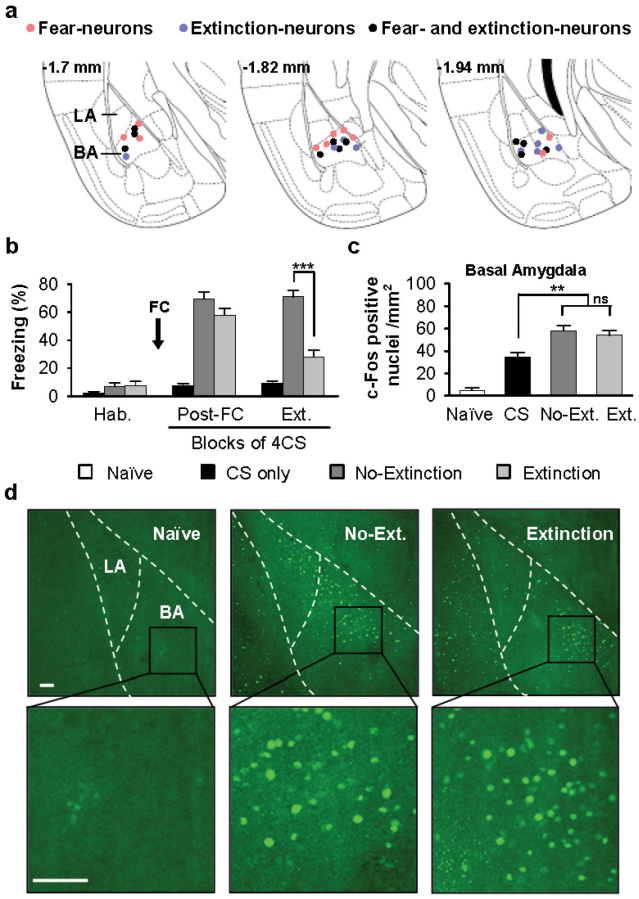

Figure 12. Fear and extinction neurons are intermingled within BA.

a, Coronal sections through the rostrocaudal extent of the amygdala showing the location of the recording wires in the BA from which activity of fear and extinction neuron was recorded. BA: basal nucleus of the amygdala; LA: lateral nucleus of the amygdala. **b**, Naïve mice (n = 7) and control animals (n = 21) exposed to the CS and to the context exhibited low freezing levels throughout the experiment. Fear conditioned animals showed high freezing levels at both time points. In mice subjected to extinction training, freezing levels were significantly reduced (Day 3, no-extinction: 71 ± 5% of time, n = 16; extinction: 28 ± 5% of time, n = 13, $P < 0.001$, two-tailed unpaired t-test). **c,** Averaged data illustrating that even though freezing behavior was significantly different, equal numbers of c-Fos expressing neurons were detected in the BA of mice exposed to an extinguished or to a non-extinguished CS (No-extinction: 58 ± 5 cells per mm2; extinction: 54 ± 4 cells per mm2, $P = 0.533$; two-tailed unpaired t-test). **d,** Examples of c-Fos expression in BA neurons of a naïve, non-extinguished and extinguished mice. $^{**}P < 0.01$, $^{***}P < 0.001$, scale bar 100 µm.

Converging evidence supports a role for the mPFC in the consolidation of extinction memory[77,104,111,120], and for the hippocampus in processing contextual information relevant for the expression and extinction of conditioned fear behavior[76]. This raises the question as to how fear- and extinction-neurons in the BA communicate with the mPFC and the hippocampus during context-dependent behavioral transitions. We first addressed the possibility that fear-neurons might be excitatory projection neurons, while extinction-neurons might be inhibitory interneurons. However, both fear- and extinction neurons exhibited low spontaneous firing rates characteristic of BLA projection neurons[121] (Table 1). Consistent with this, analysis of cross-correlations between identified fear- or extinction-neurons and neighboring BA neurons revealed short-latency excitatory interactions (Figure 13).

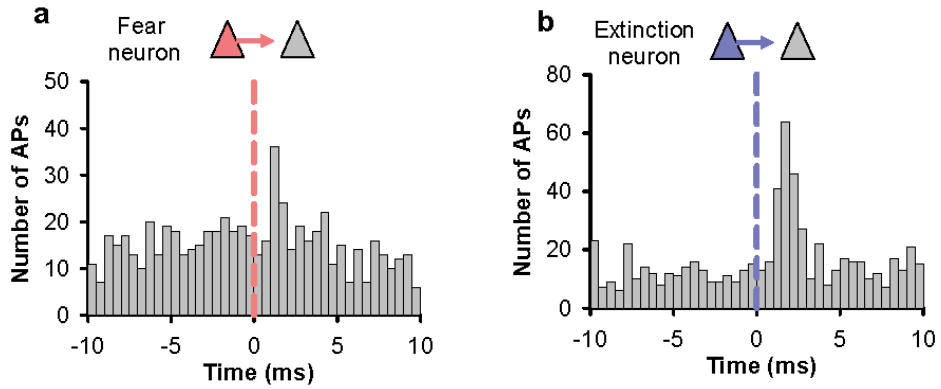

Figure 13. Cross-correlation analysis.
Consistent with the extracellular stimulation experiments, analysis of cross-correlations between identified fear- or extinction-neurons and neighboring BA neurons indicate that fear- and extinction-neurons are projection neurons. **a,** Cross-correlation between a fear-neuron and a non-identified neuron showing a short-latency, monosynaptic, excitatory interaction. Reference event is the spike of the fear neuron (dotted line at time 0). **b,** Cross-correlation between an extinction-neuron and a non-identified neuron showing a short-latency, monosynaptic, excitatory interaction. Reference event is the spike of the extinction neuron (dotted line at time 0).

To examine whether identified fear- and extinction-neurons project to, or receive input from the mPFC and/or the hippocampus, we tested for antidromic activation of BA efferents and orthodromic activation of afferents by using extracellular stimulation electrodes in re-anaesthetized mice (Figure 14a; see Methods). These experiments revealed that fear-neurons received input from the hippocampus, whereas no connections with the hippocampus inputs were found for extinction-neurons ($P < 0.05$ vs. fear-neurons; Figure 14b). While these findings cannot exclude that some extinction-neurons might be contacted by hippocampal afferents, they demonstrate that the probability of receiving hippocampal input is significantly different for fear- and extinction-neurons. Likewise, fear- and extinction-neurons were differentially connected with the mPFC. While extinction-neurons were reciprocally connected, fear-neurons projected to the mPFC, but we did not find any inputs ($P < 0.001$ vs. extinction-neurons; Figure 14b).

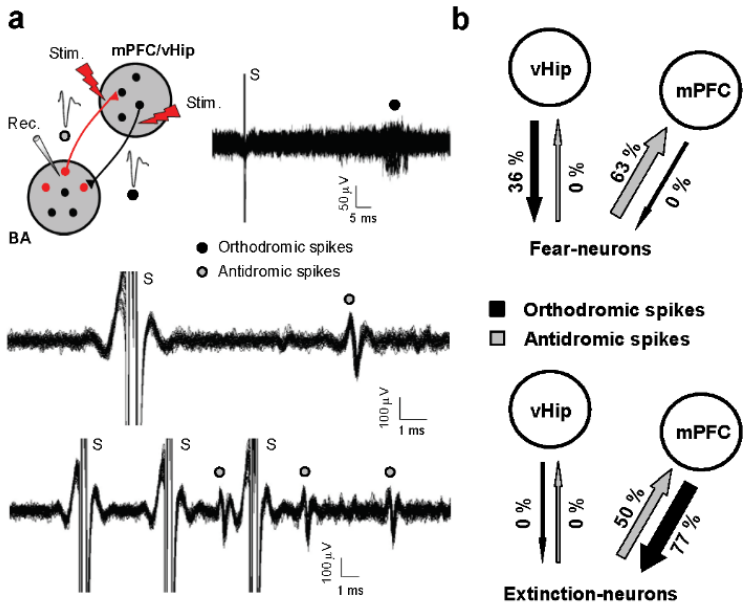

Figure 14. Fear neurons and extinction neurons are part of distinct neuronal circuits.

a, Using extracellular stimulation in anaesthetized mice to identify orthodromic and antidromic connections between BA neurons and the mPFC or the hippocampus. Top left: Schematic illustrating placement of stimulating and recording electrodes. Top right: Orthodromic spikes elicited in a BA fear-neuron upon stimulation of the ventral hippocampus. Orthodromic spikes exhibited a large temporal jitter and high failure rates. Middle: Antidromic spikes recorded from a BA extinction-neuron in response to mPFC stimulation. Antidromic spikes exhibited low temporal jitter, and followed high frequency (200 Hz) stimulation (bottom). **b,** Top: Fear-neurons project to the mPFC (5 out of 8 stimulated neurons) and receive input from the hippocampus (5 out of 14 stimulated neurons). No antidromic responses from the hippocampus (0 out of 14 stimulated neurons) or orthodormic responses from the mPFC (0 out of 8 stimulated neurons) were observed. The graph depicts the percentage of all stimulation experiments in which a particular response was observed in identified fear-neurons. Bottom: Extinction-neurons are reciprocally connected with the mPFC (antidromic responses: 3 out of 6 stimulated neurons; orthodromic responses: 7 out of 9 stimulated neurons, $P < 0.001$ vs. fear-neurons). No connections with the hippocampus were observed (0 out of 9 stimulated neurons, $P < 0.05$ vs. fear-neurons).

Extinction-resistant neurons were reciprocally connected to both the mPFC and to the hippocampus (Figure 15).

Taken together, these findings indicate that fear- and extinction-neurons, although co-localized within the same nucleus, are not only functionally specialized, but also form part of discrete neuronal circuits.

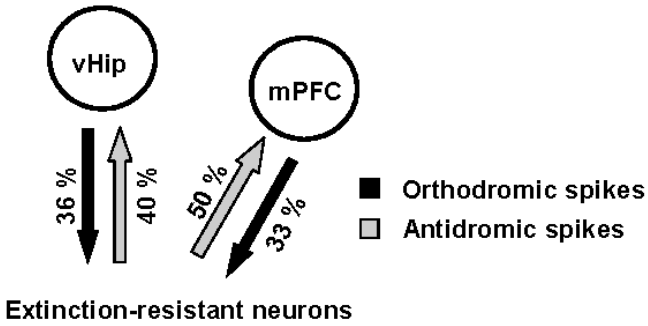

Figure 15. Connectivity of extinction-resistant neurons.

Extinction-resistant neurons are reciprocally connected to the mPFC (orthodromic responses: 3 out of 9 stimulated neurons; antidromic responses: 6 out of 12 neurons) and to the hippocampus (orthodromic responses: 4 out of 11 stimulated neurons; antidromic responses: 2 out of 5 neurons). The graph depicts the percentage of all stimulation experiments in which a particular response was observed in identified extinction-resistant neurons.

BA inactivation prevents behavioral transitions

The observed changes in CS^+-evoked spike firing of fear- and extinction-neurons during the extinction and context-dependent renewal of conditioned fear responses could be necessary for the acquisition, storage and/or behavioral expression of the learned information. To distinguish between these possibilities, we used micro-iontophoresis of a fluorescently labeled $GABA_A$ receptor agonist (muscimol) to reversibly inactivate neuronal activity in BA in a targeted and controlled manner (Figure 16a). Simultaneous iontophoresis and multi-unit recording revealed that muscimol application silenced neuronal activity in BA for more than 60 min (Figure 16b). We first tested whether BA activity was necessary for the acquisition of extinction. Inactivation of the BA completely prevented the decrease in freezing behavior normally observed during extinction training (Figure 16c), with no effect on pre-CS freezing levels (not shown). Twenty four hours later, after wash-out of muscimol, the same animals initially exhibited high freezing levels followed by normal fear extinction, demonstrating that BA inactivation did not merely interfere with the behavioral expression of extinction, nor irreversibly damage BA function (Figure 16d). These results demonstrate that BA activity is necessary for the acquisition of extinction.

Next, we tested whether BA activity was necessary for the context-dependent renewal of previously extinguished fear responses. Mice exhibiting low freezing levels during recall of extinction memory one week after extinction training were injected with muscimol before renewal. In contrast to control animals injected with the fluorescent label only, muscimol-injected animals exhibited no increase in freezing levels when placed in the fear conditioning context (Figure 16e). These results demonstrate that BA activity is necessary for context-dependent fear renewal.

Since muscimol unselectively silences all neurons in the targeted region, the high fear level observed in muscimol-injected mice during extinction learning cannot be accounted for by activity of fear neurons. Conversely, the low fear level displayed by muscimol-injected mice during context-dependent fear renewal cannot be dependent on the activation of extinction-neurons. Thus, while animals with inactivated BA are able to express high and low fear states, possibly by activation of other parts of the amygdala and the mPFC, they exhibit emotional perseveration (i.e. they remained in the emotional state they were in before BA inactivation). This suggests that the BA is unlikely to be associated with the storage, retrieval, or expression of conditioned fear and extinction

memories, but may rather mediate context-dependent behavioral transitions between low and high fear states.

Thus, silencing of BA activity should have no effect on the retrieval and expression of conditioned and extinguished fear memories when there is no need to change fear levels in a context-dependent manner. Consistent with this scenario, BA inactivation had no effect on the retrieval or expression of consolidated extinction memories (Figure 16e, Rec-2). Moreover, in animals that had been fear conditioned one week before, but that did not receive extinction training, muscimol had no effect on the retrieval and expression of the fear memory independently of the context in which they were tested (Figure 16f).

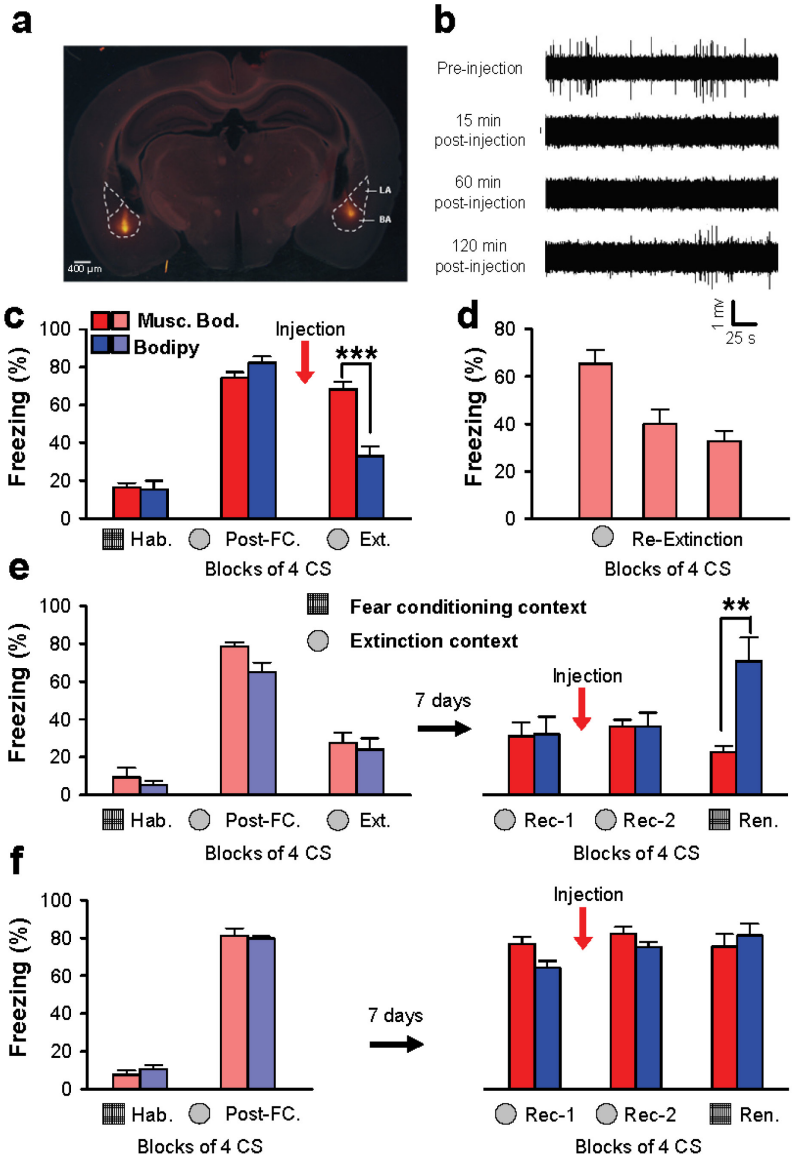

Figure 16. Targeted inactivation of the BA prevents behavioral changes without affecting memory. **a,** Epifluorescent image illustrating bilateral targeting of the BA with fluorescently labeled muscimol (muscimol-bodipy). **b,** Simultaneous multi-unit recordings reveal silencing of neuronal activity for up to two hours after muscimol iontophoresis. **c,** Inactivation of the BA before extinction training prevents the acquisition of extinction. Control mice injected with fluorophore only (n = 5) exhibited significant reduction of freezing levels after extinction training. Muscimol-injected animals (n = 11) showed high freezing levels after extinction. **d,** Twenty four hours later, in the absence of muscimol, the same animals showed normal acquisition of extinction (P < 0.05). **e,** Inactivation of the BA prevents context-dependent renewal. Control mice injected with fluorophore only (n = 5) exhibited a significant increase in freezing levels upon change of context (P < 0.05). Muscimol-injected animals (n = 5) do not show any context-dependent fear renewal (P < 0.01 vs. control). **f,** In the absence of extinction training, BA inactivation did not affect fear memory retrieval. Fluorophore-injected mice (n = 4) and muscimol-injected mice (n = 5) exhibited equal freezing levels during CS^+ exposure in the fear conditioning context one week after fear conditioning. $^{**}P$ < 0.01, $^{***}P$ < 0.001.

Discussion

Our data shows that the BA contains distinct populations of neurons whose activity is oppositely correlated with high and low fear behavior, two converse behavioral states. Although fear- and extinction-neurons represent relatively small sub-populations within the BA, a rapid switch in the balance of their activity is essential for triggering behavioral transitions during extinction and context-dependent fear renewal. While intermingled within the BA, fear- and extinction-neurons are differentially connected with the hippocampus and the mPFC, two brain areas previously implicated in extinction and context-dependent renewal of conditioned fear responses. In keeping with the proposed role of the ventral hippocampus in mediating context-dependent renewal of fear behavior in animals subjected to extinction[114], we found that hippocampal input to the BA selectively targets fear-neurons over extinction-neurons. Thus, hippocampal input to BA fear-neurons may override the retrieval of extinction memory allowing for fear expression after a particular CS has undergone extinction. Extinction-neurons, in turn, are bi-directionally connected with the mPFC and are switched on during extinction training. This indicates that they may be upstream of a previously identified population of mPFC neurons thought to mediate consolidation of extinction memory, because they are activated by an extinguished CS during recall, but not during the acquisition of extinction[104].

Previous findings demonstrate that the BLA is not critical for triggering behavioral transitions during reversal learning in a two odor discrimination task[122,123]. Nevertheless, abnormally persistent BLA activity induced by orbitofrontal cortex lesions[122] or repeated cocaine administration[123] interferes with reversal learning. This suggests that while the BLA can veto slow behavioral transitions during more complex reversal learning tasks, it is actively involved in situations requiring rapid context-dependent switching between two converse behavioral states.

How might activity of BA fear- and extinction-neurons mediate behavioral transitions? In keeping with a role for the amygdala in facilitating network function and memory formation in other parts of the brain[124–126], a possible interpretation is that BA fear- and extinction-neurons might drive or facilitate the induction of synaptic plasticity in their respective target areas. Moreover, while previous studies using pre-fear conditioning lesions came to the conclusion that the BA does not contribute to the acquisition or the expression of conditioned fear[46,127–129] (but see ref. 130), a recent analysis using post-fear conditioning lesions indicates that the BA also contributes to the

consolidation of long-term fear memories[129]. This suggests that repeated activity of BA fear-neurons, over longer-time periods, may be required for fear memory consolidation.

Our findings are consistent with the idea that in mammals, as in invertebrates[131,132] switches between appropriate behavioral states can be driven by discrete neuronal circuits. Although it remains to be shown how fear- and extinction-neurons interact locally, it may be a general principle of the functional micro-architecture of the nervous system in diverse species that circuits mediating switches between distinct behavioral states are located in close anatomical proximity thereby allowing for local interactions. Finally, our results also suggest that context-dependent recovery of extinguished fear behavior in humans[133], which represents a major clinical obstacle for the therapy of certain anxiety disorders[134], might be modulated by tipping the balance of activity between specific neuronal circuits.

Material and methods

Animals

Male C57BL6/J mice (3 months old; RCC Ltd., Füllinsdorf, Switzerland) were individually housed for 7 days prior to all experiments, under a 12 h light/dark cycle, and provided with food and water *ad libitum*. All animal procedures were executed in accordance with institutional guidelines and were approved by the Veterinary Department of the Canton of Basel-Stadt.

Behavior

Fear conditioning and extinction took place in two different contexts (Context A and B). The conditioning and extinction boxes and the floor were cleaned with 70% ethanol or 1% acetic acid before and after each session, respectively. To score freezing behavior an automatic infrared beam detection system placed on the bottom of the experimental chambers (Coulbourn Instruments, Allentown, PA) was used. The animals were considered to be freezing if no movement was detected for 2 s. On day 1, mice were submitted to a habituation session in context A, in which they received 4 presentations of the CS^+ and the CS^- (total CS duration: 30 s, consisting of 50 ms pips repeated at 0.9 Hz, 2 ms rise and fall, pip frequency: 7.5 kHz or 3 kHz, 80 dB). Discriminative fear conditioning was performed the same day by pairing the CS^+ with a US (1 s foot-shock, 0.6 mA, 5 CS^+-US pairings; inter-trial interval: 20-180 s). The onset of the US coincided with the offset of

the CS^+. The CS^- was presented after each CS^+/US association but was never reinforced (5 CS^- presentations, inter-trial interval: 20-180 s). The frequencies used for CS^+ and CS^- were counterbalanced across animals. On day 2 and day 3, conditioned mice were submitted to extinction training in context B during which they received 4 and 12 presentations of the CS^- and the CS^+, respectively. Recall of extinction and context-dependent fear renewal were tested 7 days later in context B and A, respectively, with 4 presentations of the CS^- and the CS^+. Pharmacological experiments were performed using the same conditioning and extinction protocol except for one group of mice that was not submitted to extinction training but tested for conditioned fear with 4 CS^- and 4 CS^+ presentations on day 2 in context B. Seven days later, mice were submitted to 2 sessions of extinction recall 5 h apart in context B (4 presentations of each CS for each session). Finally, 10 min after the second recall session, mice were submitted to 4 CS^- and 4 CS^+ presentations in context A for context-dependent fear renewal.

For discriminative extinction, mice were habituated on day 1 to 4 presentations of two different CSs in context A (total CS duration: 30 s, consisting of 50 ms pips repeated at 0.9 Hz, 2 ms rise and fall, pip frequency: 7.5 kHz or 3 kHz, 80 dB). Both CSs were subsequently paired with a US (1 s footshock, 0.6 mA, 5 CS/US pairings for each CS; inter-trial interval: 20-180 s). The onset of the US coincided with the offset of the CSs. On day 3 and 4, only one of the two CSs was extinguished by 16 and 12 presentations in context B, respectively. At the end of the second extinction session, mice were exposed to 4 presentations of the non-extinguished CS in context B.

Surgery and recordings

Mice were anesthetized with isoflurane (induction 5%, maintenance 2.5%) in O_2. Body temperature was maintained with a heating pad (CMA/150, CMA/Microdialysis, Stockholm, Sweden). Mice were secured in a stereotaxic frame and unilaterally implanted in the amygdala with a multi-wire electrode aimed at the following coordinates[135]: 1.7 mm posterior to bregma; ± 3.1 mm lateral to midline and 4 to 4.3 mm deep from the cortical surface. The electrodes consisted of 8 to 16 individually insulated nichrome wires (13 µm inner diameter, impedance 1-3 MΩ; California Fine Wire, Grover Beach, CA) contained in a 26 gauge stainless steel guide canula. The wires were attached to a 10 to 18 pin connector (Omnetics, Minneapolis, MN). The implant was secured using cyanoacrylate adhesive gel. After surgery mice were allowed to recover for 7 days.

Analgesia was applied before, and during 3 days after surgery (Metacam, Boehringer, Basel, Switzerland). Electrodes were connected to a headstage (Plexon, Dallas, TX) containing eight to sixteen unity-gain operational amplifiers. The headstage was connected to a 16-channel computer controlled preamplifier (gain 100x, bandpass filter from 150 Hz to 9 kHz, Plexon). Neuronal activity was digitized at 40 kHz and bandpass filtered from 250 Hz to 8 kHz, and isolated by time-amplitude window discrimination and template matching using a Multichannel Acquisition Processor system (Plexon). At the conclusion of the experiment, recording sites were marked with electrolytic lesions before perfusion, and electrode locations were reconstructed with standard histological techniques.

Single-unit spike sorting and analysis

Single–unit spike sorting was performed using Off-Line Spike Sorter (OFSS, Plexon) as described[136,137] (Figure 17). Principal component (PC) scores were calculated for unsorted waveforms and plotted on 3D PC spaces and clusters containing similar valid waveforms were manually defined. A group of waveforms was considered to be generated from a single neuron if it defined a discrete cluster in PC space that was distinct from clusters for other units and if it displays a clear refractory period (> 1 ms) in the auto-correlogram histograms. In addition, two parameters were used to quantify the overall separation between identified clusters in a particular channel. These parameters include the J3 statistic that corresponds to the ratio of between-cluster to within-cluster scatter, and the Davies-Bouldin validity index (DB) that reflects the ratio of the sum of within-cluster scatter to between-cluster separation[137]. High values for the J3 and low values for the DB are indicative of good single unit isolation (Figure 17). Controls values for this statistics were obtained by artificially defining two clusters from the centered cloud of points in the PC space from channels in which no units could be detected. Template waveforms were then calculated for well separated clusters and stored for further analysis. Clusters of identified neurons were analyzed offline for each recording session using principal component analysis and a template matching algorithm. Only stable clusters of single units recorded over the time course of the entire behavioral training were considered. Long-term single unit stability isolation was first evaluated using Wavetracker (Plexon) in which PC space-cylinders were calculated from 5 min segment of data spontaneously recorded before any training session. Straight cylinders suggest that the same

set of single units was recorded during the entire training session (Figure 17). Secondly, we quantitatively evaluated the similarity of waveform shape by calculating linear correlations (*r*) values between average waveforms obtained over training days[138] (Figure 17).

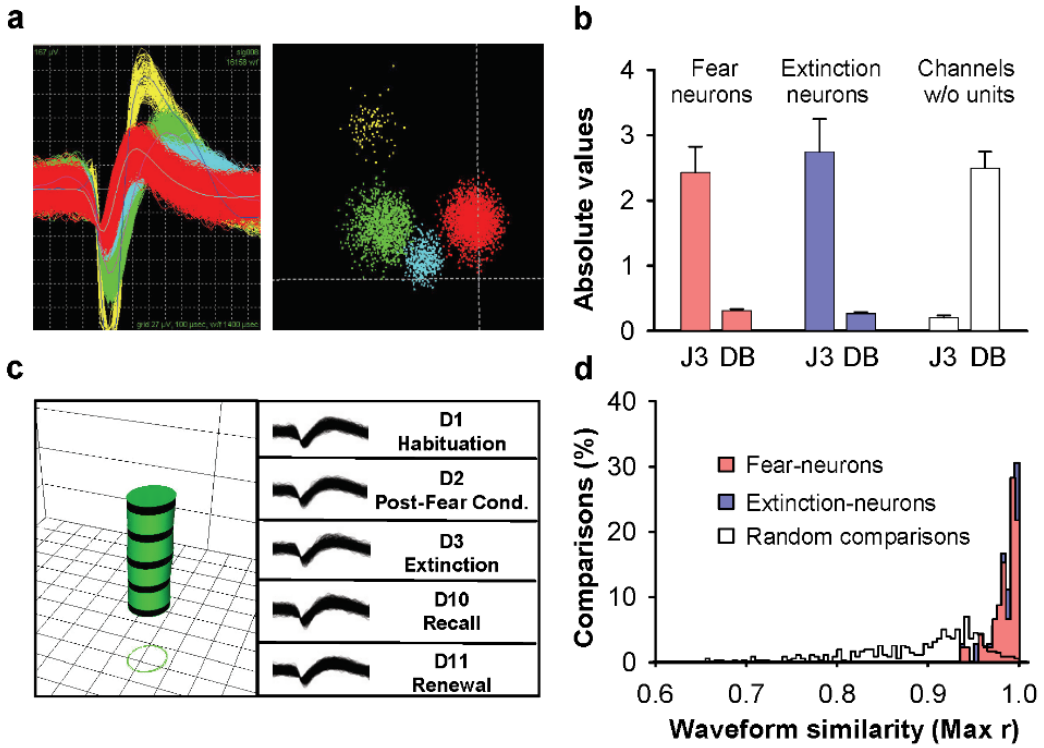

Figure 17. Stability of chronic single unit recordings from mouse amygdala.
a, Top left: Superimposed waveforms recorded from four different units. Top right: Spikes originating from individual units were sorted using 3D-principal component analysis. **b,** Quantitative J3 and Davies Bouldin validity index (DB) statistics calculated for fear and extinction neurons. Controls values were obtained using two clusters defined from the centered cloud of points from channels in which no units could be detected. High values for the J3 and low values for the DB are indicative of good single unit isolation. **c,** Left**:** Stability of clustered waveforms during long-term recordings was assessed by calculating principal component (PC) space cylinders. Straight cylinders suggest that the same set of single units was recorded during the entire training session. Right: Superimposed waveforms used to calculate the PC space cylinder recorded before habituation, extinction, recall and renewal sessions**. d,** In addition, to quantitatively evaluate similarity of different spike shapes recorded on different days, linear correlation values between time-shifted average waveforms were calculated for fear and extinction neurons. As a control we computed the *r* values from average waveforms of different neurons. The maximum *r* value across time shifts was used to quantify similarity ($r = 1$ would indicates identical spike shapes). These calculations revealed that 94.4% of extinction neurons and 95.65 % of fear neurons had an *r* value above 0.95, compared with only 17.9% of similarity scores calculated between waveforms of different cells.

As a control we computed the *r* values from average waveforms of different neurons. Thirdly, for each unit we used correlation analysis to quantitatively compare similarity of waveform shape during CS$^+$-stimulation and during a 60 s period of spontaneous activity recorded prior to each behavioral session. (Figure 18).

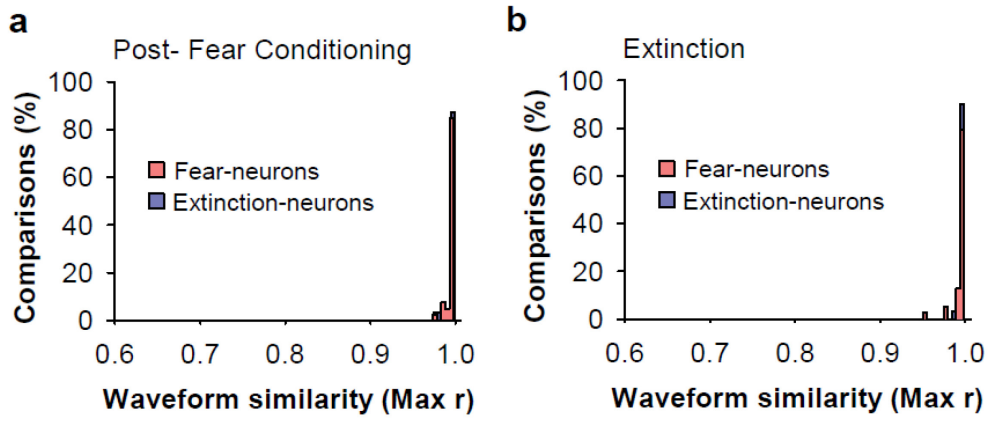

Figure 18. Quantitative comparisons of waveforms across periods of spontaneous activity and sensory stimulation.
a, For each identified fear- and extinction-neuron we calculated linear correlation values between time-shifted average waveforms obtained during a 60 s period of spontaneous activity recorded prior to each behavioral session and during CS stimulation. The maximum r value across time shifts was used to quantify similarity (r = 1 would indicates identical spike shapes). These calculations revealed r values above 0.95 for 100% of all units. **b,** Same plot for all units recoded before and during the extinction session.

To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at a time that the reference neuron fires, only one of the two neurons was considered for further analysis. CS-induced neural activity was calculated by comparing the firing rate after stimulus onset with the firing rate recorded during the 500 ms before stimulus onset (bin size: 20 ms; averaged over blocks of 4 CS presentations consisting of 108 individual sound pips in total) using a *z*-score transformation. *Z*-score values were calculated by subtracting the average baseline firing rate established over the 500 ms preceding stimulus onset from individual raw values and by dividing the difference by the baseline standard deviation. Only CS-excited neurons were considered for analysis. Classification of units was performed by comparing the largest significant *z*-score values within 100 ms following CS-onset during post-fear conditioning and extinction sessions according to the freezing levels. For high fear states, the entire post-fear conditioning session was analyzed, whereas for low fear states, analysis was restricted to the block of 4 CS presentations during which the fear level was the

lowest. A unit was classified as a fear-neuron if it exhibited a significant *z*-score value after fear conditioning (when freezing levels were high), but no significant *z*-score value after extinction (when freezing levels were low), and vice versa for extinction-neurons. Finally, units were classified as extinction-resistant neurons if they displayed a significant *z*-score value during both post-fear conditioning and extinction sessions, independently of freezing levels. For statistical analysis, *z*-score comparisons were performed using the average *z*-score value calculated during the 40 ms following CS-onset. In cases where shorter or longer CS-evoked activity was observed, the average *z*-score was calculated during the 20 ms and 80 ms following CS-onset, respectively. To identify the trial in which individual neurons changed their CS-evoked responses during fear conditioning and extinction, we applied a change point analysis algorithm[119]. Change point analysis identifies the trial(s) exhibiting a significant change in neuronal activity or freezing behavior relative to the preceding trials. Change points are graphically represented by a change in the slope of a plot showing the cumulative sums of the averaged and normalized *z*-score and freezing values. Statistical analyses were performed using paired Student's *t*-tests *post hoc* comparisons at the $P < 0.05$ level of significance unless indicated otherwise. Results are presented as mean ± S.E.M.

Extracellular stimulation

In order to determine the connectivity of recorded neurons, we used extracellular stimulation of the medial prefrontal cortex (mPFC) and the ventral hippocampus (vHip) in a subset of animals. At the end of the training procedure, animal were anesthetized using urethane (1.4 g kg^{-1}) and concentric stimulating electrodes (FHC, Bowdoin, ME) were lowered in the mPFC (2 mm anterior to bregma; ± 0.3 mm lateral to midline and 1.6 to 2 mm deep from the cortical surface) and the ventral hippocampus (3.6 mm posterior to bregma; ± 3.1 mm lateral to midline and 4 to 4.2 mm deep from the cortical surface). During the experiments, the stimulation electrodes were advanced in steps of 5 μm by a motorized micromanipulator (David Kopf Instruments, Kujunga, CA) and BA-evoked responses were recorded. Stimulation-induced and spontaneous spikes were sorted using principal component analysis and template matching. Similarity of stimulation-induced spike waveforms was quantitatively compared to the waveforms of units previously

identified in the awake animal and recorded on the same wire using correlation analysis (Figure 19).

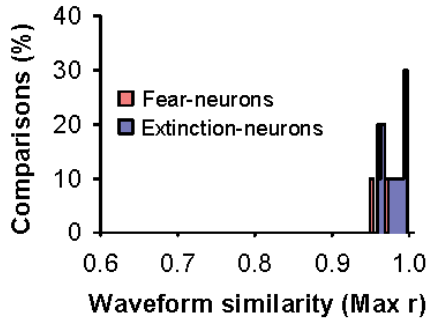

Figure 19. Identification of units activated by extracellular stimulation. Similarity of stimulation-induced spike waveforms was quantitatively compared to the waveforms of fear- and extinction-neurons previously identified in the awake animal and recorded on the same wire using correlation analysis. For each unit we calculated linear correlation values between time-shifted average waveforms obtained during the extinction session and during extracellular stimulation in the anaesthetized animal. The maximum *r* value across time shifts was used to quantify similarity ($r = 1$ would indicates identical spike shapes). These calculations revealed r values above 0.95 for 100% of all units.

To be classified as antidromic, evoked-responses had to meet at least two out of three criteria: (1) stable latency (< 0.3 ms jitter), (2) collision with spontaneously occurring spikes, and (3) ability to follow high-frequency stimulation (200 Hz). At the end of the experiments, stimulating sites were marked with electrolytic lesions before perfusion, and electrode locations were reconstructed with standard histological techniques. For each stimulation site orthodromic and antidromic response probabilities of fear- and extinction-neurons were analyzed using binomial statistics with $P < 0.05$ indicating non-random connectivity.

Muscimol iontophoresis

Muscimol micro-iontophoresis injection was performed in chronically implanted animals. Single barrel micropipettes with a tip diameter of 10 to 15 µm were cut at 1 cm length and filled with a solution containing muscimol covalently coupled to a fluorophore (Muscimol-Bodipy-TMR conjugated, Invitrogen, Rockville, MD)(5 mM in phosphate buffered saline (PBS) 0.1 M, DMSO 40%) or with Bodipy alone (Invitrogen; 5mM in phosphate buffered saline (PBS) 0.1 M, DMSO 40%). Mice were bilaterally implanted at the following coordinates (according to Franklin and Paxinos, 1997)[135]: 1.7 mm posterior to bregma; ± 3.1 mm lateral to midline and 4 to 4.3 mm deep from the cortical surface. Chlorided silver wires were inserted in each micropipette and attached to a connector. A third silver wire screwed onto the skull and attached to the connector served as a reference electrode. The entire miniature was secured using cyanoacrylate adhesive gel. After surgery, mice were allowed to recover for 2 days. On the injection day, iontophoretic applications were performed by means of cationic current (+12 to +15 nA) for 15 min per side using a precision current source device (Stoelting, Kiel, WI). Mice were submitted to the behavioral procedure 5

min after the end of iontophoretic injections and immediately perfused at the end of the experiments. Brains were collected for further histological analysis. Serial slices containing the amygdala were imaged at 5X using an epifluorescence stereo microscope (Leica, Wetzlar, Germany), and the location and the extent of the injections were controlled. Mice were included in the analysis only if they presented a bilateral injection targeting exclusively the BA and if the targeted injections cover at least 25% of the BA. Statistical analyses were performed using paired and unpaired Student's *t*-tests *post hoc* comparisons at the $P < 0.05$ level of significance. Results are presented as mean ± S.E.M..

Behavior and pharmacological inactivations

Mice were submitted to a discriminative auditory fear conditioning paradigm in which the CS^+, but not the CS^-, was paired with a US (mild foot-shock). Extinction training was performed over two days in a different context[109]. One week later, mice were placed in the extinction context for recall of extinction, and in the original conditioning context for fear renewal. Freezing behavior was quantified during each behavioral session using an automatic infrared beam detection system as previously described[136]. Bilateral inactivation of the BA was achieved using micro-iontophoretic injection of fluorescently labeled muscimol before extinction training or context-dependent fear renewal.

Electrophysiological recordings and analysis.

Individual neurons were recorded extracellularly in freely behaving mice during fear conditioning, extinction, recall of extinction and context-dependent fear renewal. Spikes of individual neurons were sorted by time-amplitude window discrimination and template matching as previously described[136,137]. Cluster quality and unit stability was verified by quantifying the cluster separation and the stability of the average waveform shape over time[137,138] (Figure 17). Unit isolation was verified using auto- and cross-correlation histograms. Spike rasters and histograms were constructed by aligning sweeps relative to the CS onset, and CS-evoked responses were normalized to baseline activity using a *z*-score transformation. Antidromic and orthodromic spikes evoked by extracellular stimulations of the mPFC or the ventral hippocampus were recorded in previously identified neurons in anaesthetized mice.

Immunohistochemistry

Mice were transcardially perfused with ice-cold 4% paraformaldehyde in 0.1 M phosphate buffered saline 120 min after the onset of the training session[116]. Brains were prepared for immunohistochemistry using primary polyclonal rabbit anti-c-Fos antibody (Calbiochem, San Diego, CA)(Anti-c-Fos, Ab-5, 4-17, Rabbit pAb, PC38; 1:20000 dilution). A fluorescent dye-coupled goat anti-rabbit antibody (Invitrogen; Alexa-Fluor 633; 1:1000 in PBS) was used as secondary antibody. Stained slices were imaged at 40X using a LSM 510 Meta confocal microscope (Carl Zeiss Inc., Jena, Germany). Quantitative analysis of c-Fos-positive nuclei was performed using a computerized image analysis system (Imaris 4.2, Bitplane, Zürich, Switzerland). Structures were defined according to Franklin and Paxinos (1997)[135]. Immunoreactive neurons were counted bilaterally using a minimum of three sections per hemisphere per animal. Statistical analyses were performed using unpaired Student's *t*-tests at the $P < 0.05$ level of significance. Results are presented as mean ± S.E.M..

INTERACTION BETWEEN CONDITIONING AND EXTINCTION OF OPPOSITE VALENCES

Results

Pavlovian appetitive conditioning in mice

In order to be able to compare the neuronal activity elicited by aversive and appetitive cues, an appetitive conditioning paradigm was developed to assess appetitive associative learning in mice conditioned in a purely Pavlovian manner. In this Pavlovian appetitive conditioning, as for classical FC, the US is passively received by the mice: no approach or instrumental behavior is required for the animal to actually obtain a reward. To achieve this, a surgical procedure for the implantation of an intra-oral cannula in mice was implemented allowing for the delivery of a sucrose solution directly into the oral cavity upon CS presentation.

Description of the appetitive learning procedures is detailed in the material and methods sections of this dissertation. Briefly, mice were first exposed to a CS^- and to a CS^+_{ap} during the habituation session. On the next day, they were submitted to Pavlovian appetitive conditioning during which only the CS^+_{ap} was paired with an intra-oral delivery of the sucrose solution. Finally, on the following day, mice underwent an appetitive extinction session during which they were exposed to the CS^- and to unreinforced presentations of the CS^+_{ap} (Figure 20a).

During US_{ap} delivery, mice showed typical hedonic taste reactivity (set of orofacial movements previously described in the Rats as reflective of the palatability of tastants[139]). Consistent with Pavlov's substitution theory[16], which postulates that URs gradually transfer from the US to the CS, appetitive CRs identical to the URs were progressively expressed during the appetitive conditioning at the time of the CS^+_{ap}. Assessing the emotional significance gained by the CS^+_{ap} upon appetitive conditioning was achieved by comparing the orofacial taste reactivity before *versus* after the appetitive conditioning. In order to do that, orofacial movements were further classified into three different types of actions depending on their valence: hedonic (i.e. paw lickings, tongue protrusions and licking/eating of items in the context), aversive (i.e. head flails and gapes) and neutral orofacial movements (i.e. mouth openings and lateral chin movements). The comparison of the number of orofacial movements before and after conditioning shows that appetitive conditioning leads to an increase in orofacial movements, specifically in hedonic ones. No aversive orofacial movements and only a few neutral orofacial movements are expressed before and after

appetitive conditioning, while the expression of hedonic orofacial movements drastically increases after appetitive conditioning (Figure 20b).

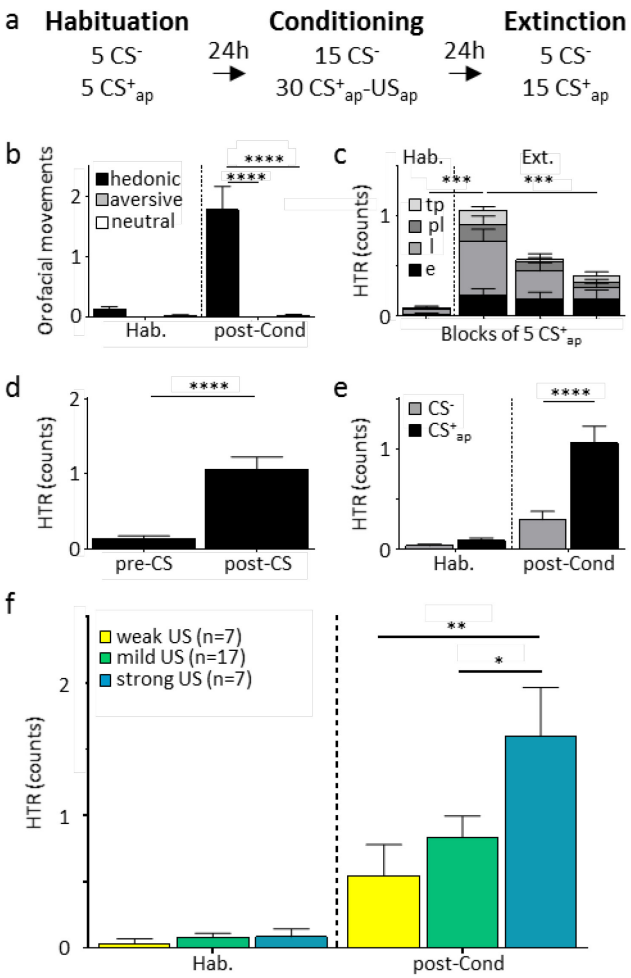

Figure 20. Pavlovian appetitive conditioning
a: Behavioral protocol. **b**: Orofacial movements (counts) before and after Pavlovian appetitive conditioning. Hab., before appetitive conditioning; post-Cond, after appetitive conditioning. Two-way ANOVA, Sidak correction for multiple comparisons, p<0.0001. **c**: Individual components of hedonic taste reactivity (HTR), before and after Pavlovian appetitive conditioning. Hab., Habituation; Ext., appetitive extinction; tp, tongue protrusions; pl, paw licking; l, licking of items in the arena; e, eating items in the arena. Repeated-measure one-way ANOVA, Tukey correction for multiple comparisons, p<0.0001. **d**: Hedonic taste reactivity after Pavlovian appetitive conditioning, during the ten seconds prior to cue presentation (pre-CS) and during CS^{+}_{ap} exposure (CS). Two-tailed paired t-test, p<0.0001. **e**: Hedonic taste reactivity evoked by CS^{-} and CS^{+}_{ap} presentations before (Hab.) and after Pavlovian appetitive conditioning (post-Cond). Two-way ANOVA, Sidak correction for multiple comparisons, p<0.001. **f**: Hedonic taste reactivity to CS^{+}_{ap} presentations before (Hab.) and after (post-Cond) Pavlovian appetitive conditionings of different intensities. Weak US, 20 µL of 0.05 M sucrose solution; mild US, 20 µL of 0.8 M sucrose solution; strong US, 20 µL of 1 M sucrose solution. Two-way ANOVA, Sidak correction for multiple comparisons, p<0.05. N= 24 mice (from mild and strong Pavlovian appetitive conditioning, unless otherwise specified on the graph); Error bars indicate mean ± s.e.m. *: p<0.05; **: p<0.01; ***: p<0.001; ****: p<0.0001.

Hedonic taste reactivity (HTR), is a composite index consisting of weighted hedonic orofacial movements (see the Material and Methods section for more details on the weighting) and has been used to assess the palatability of tastants[139–141]. Our results demonstrate that HTR can be used in mice as an appetitive conditioned response and that HTR reflects the learning of Pavlovian appetitive conditioning and extinction. Indeed, HTR increases after appetitive conditioning and gradually decreases upon repetitive presentations of the CS^{+}_{ap} alone (Figure 20c). Importantly, mice do not express HTR outside of the cue period: CR_{ap} are evoked by the CS presentation and almost no HTR is detected during the period preceding the CS^{+}_{ap} (Figure 20d). This also indicates that appetitive conditioning does not lead to a general increase in hedonic orofacial movements but rather to a specific expression of CR_{ap} locked to the cue presentation. Additionally, conditioned

mice show a very clear behavioral discrimination between the cue which was associated with the US_{ap} and the non-reinforced cue, expressing HTR after conditioning only during the CS^{+}_{ap} and not during the presentation of the CS^{-} (Figure 20e). These observations indicate that only the cue which was paired with the US_{ap} has gained an emotional significance and suggest that HTR reflects the learned emotional significance carried by the CS^{+}_{ap} after appetitive conditioning. Finally, to confirm that HTR can be used to assess the learned hedonic significance of reward-predicting cues in mice, post-conditioning HTR levels were compared between groups receiving US_{ap} of different intensities. Increased concentrations of the sucrose solution used as a US_{ap} induce higher levels of CR_{ap}, confirming that HTR represents the relative palatability of reward-associated cues (Figure 20f). Taken together these results demonstrate that specific types of orofacial movements, the HTR, can be used, in mice, to infer the hedonic significance of CSs and thus assess Pavlovian appetitive conditioning.

Amygdala encoding of Pavlovian appetitive conditioning and extinction

To investigate the neuronal correlates of Pavlovian appetitive conditioning, single unit recordings were performed while the mice were subjected to appetitive excitatory and inhibitory learning. Mice implanted with electrodes in the basal nucleus of the amygdala (BA) were first submitted to a habituation session, followed at a 24h interval by a Pavlovian appetitive conditioning and by an appetitive extinction session on the next day (Figure 21a). As described above, mice show an increase in HTR level as a result of Pavlovian appetitive conditioning and a decrease of their behavioral responses across appetitive extinction learning. As similarly described in the previous paragraph, this change in behavior is specifically elicited by the cue associated with the US_{ap}: CR_{ap} are significantly elevated during the first post-conditioning block of CS^{+}_{ap} presentations, but neither to the context (BL: baseline) nor to the CS^{-} (Figure 21b). This valence-specific behavioral response decreases through appetitive extinction training until it reaches a level similar to the pre-conditioning one (Figure 21b). These results indicate that mice which underwent single unit recordings in the BA have indeed learned about the different contingencies between the CS^{+}_{ap} and the US_{ap} both during the excitatory and inhibitory phases of appetitive learning.

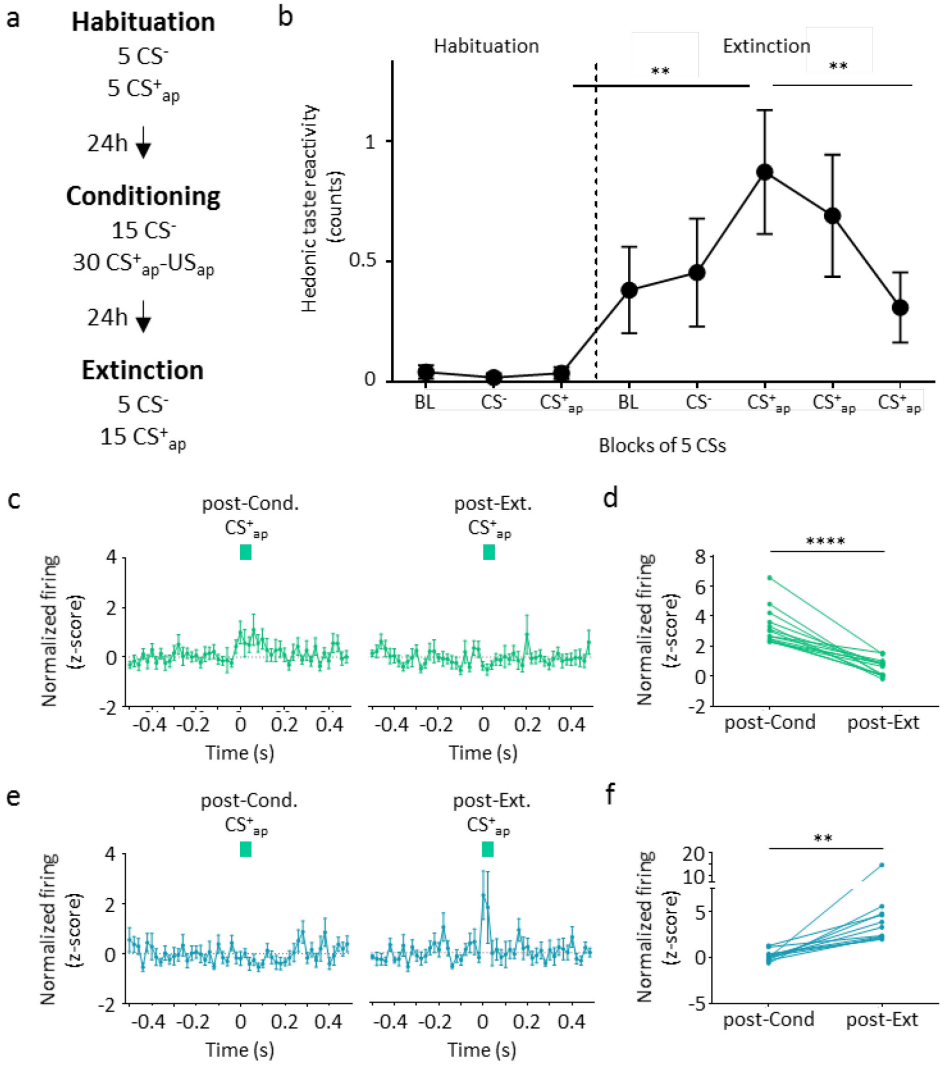

Figure 21. Neuronal correlates of Pavlovian appetitive conditioning and extinction

a: Behavioral protocol. **b**: Behavioral performance before (Habituation) and after (Extinction) Pavlovian appetitive conditioning in amygdala-implanted mice (n=11). BL, baseline (context exposure). Animals show no HTR (Hedonic taste reactivity) prior to conditioning and exhibit a specific increase in HTR during post-conditioning CS^{+}_{ap} presentations which is reversed by appetitive extinction training. Repeated-measure one-way ANOVA, uncorrected Fisher's LSD test, $p<0.05$. **c**: Normalized activity (z-score) of appetitive neurons after Pavlovian appetitive conditioning (post-Cond) and at the end of extinction training (post-Ext.). Appetitive neurons (n=14) are specifically excited by CS^{+}_{ap} presentations during high but not during low hedonic state. **d**: Peak of normalized activity for individual appetitive neurons during high (post-Cond) and low (post-Ext) hedonic states. Appetitive neurons show a significant decrease of their CS^{+}_{ap}-responsiveness upon appetitive extinction learning. Two-tailed paired t-test, $p<0.0001$. **e**: Normalized activity (z-score) of appetitive extinction neurons after Pavlovian appetitive conditioning (post-Cond) and at the end of extinction training (post-Ext.). Appetitive extinction neurons (n=11) are specifically excited by extinguished appetitive cues. **f**: Peak of normalized activity for individual appetitive extinction neurons during high (post-Cond) and low (post-Ext) hedonic states. Appetitive extinction neurons show a significant increase of their CS^{+}_{ap}-responsiveness upon appetitive extinction learning. Two-tailed paired t-test, $p<0.01$. Error bars indicate mean ± s.e.m.; **: $p<0.01$; ****: $p<0.0001$.

To identify the neuronal correlates of appetitive conditioning and extinction under purely Pavlovian settings, chronic single unit recordings were performed in the BA. Two distinct neuronal subpopulations responding to appetitively conditioned cues were identified, activated specifically

either during the high or the low hedonic states of the animals. When the appetitive memory is retrieved by the CS^{+}_{ap} presentations during the first block of appetitive extinction, appetitive neurons show a phasic increase of their normalized activity (measured by z-score). Importantly, these appetitive neurons do not exhibit any phasic excitation during the presentation of the extinguished CS^{+}_{ap} (Figure 21c). This specific pattern of neuronal activity suggests that the CS^{+}_{ap}-induced responsiveness of appetitive neurons is specific to the high hedonic valence of the cue. The comparison of the peak activity of appetitive neurons post-conditioning with the post-extinction CS^{+}_{ap}-responsiveness shows a significant decrease of the cue-induced normalized firing of these neurons (Figure 21d), confirming that individual appetitive neurons show extinction-induced plasticity as a result of inhibitory appetitive learning. These results could suggest that appetitive neurons contribute to the appetitive memory trace. Conversely, appetitive extinction neurons show the exact opposite pattern of cue-responsiveness: they are not responsive to the CS^{+}_{ap} presentations after appetitive conditioning but have a phasic increase in their firing during the presentation of extinguished appetitive cues (Figure 21e). Opposite to appetitive neurons, this neuronal population increases its CS^{+}_{ap}-induced peak activity upon appetitive extinction learning (Figure 21f). This specific excitation of appetitive extinction neurons in response to appetitive extinguished cues suggests that this neuronal population is involved in the appetitive extinction memory trace. The identification of these two distinct populations of BA neurons specifically responding to either appetitive cues or to extinguished cues demonstrates that appetitive conditioning and extinction recruit two different amygdala circuits. It suggests that the retrieval of appetitive conditioning and appetitive extinction rely on two separate sets of neurons which might be responsible for the expression of appetitive memory and appetitive extinction memory in a context-dependent manner.

Amygdala encoding of Pavlovian fear conditioning and extinction

To investigate the neuronal encoding of fear and fear extinction, chronic single unit activity in the BA was recorded while mice underwent classical fear conditioning and extinction. Briefly, following habituation session to two auditory cues in absence of any reinforcement, mice were submitted to CS^{+}_{av} and US_{av} pairings. During the two consecutive days, the mice were subjected

to two fear extinction sessions (Figure 22a) (see Material and Methods section for detailed description of fear learning behavioral procedures).

Freezing behavior, a well-establish read-out of fear conditioning in Rodents, was used to monitor fear learning and fear extinction during these different sessions. Prior to fear conditioning, mice freely explore the arena and do not show freezing behavior when exposed to either one of the auditory cues (Figure 22b). Fear conditioning induces a significant increase in the percentage of time spent freezing during the CS^{+}_{av}. Post-conditioning freezing levels to the context (BL: baseline) or to the CS^{-} remain similar to pre-conditioning levels, indicating that the behavioral fear response is specific of the cue associated with the foot-shock and that mice have learned the specific emotional significance of context and of the two auditory cues (Figure 22b). Two consecutive days of fear extinction training lead to a progressive decline in the freezing response to the CS^{+}_{av}, finally reaching pre-conditioning levels when the animals learned that the CS^{+}_{av} does not predict anymore the US_{av} delivery (Figure 22b). This behavioral data shows that mice implanted for recordings of the BA neuronal activity learned discriminative fear conditioning and fear extinction, thus allowing for the investigation of the encoding of fear and fear extinction in the BA at the single cell level.

As previously described in the first part of this dissertation, the BA circuits encompass two distinct neuronal populations specifically responding to either fear or fear-extinguished cues[63].

Likewise, we identified fear neurons which are specifically activated by the presentation of the CS^{+}_{av} after FC (when the animals are in a high fear state) but not after fear extinction when presentations of the CS^{+}_{av} do not lead anymore to fear expression (Figure 22c). The peak activity of the fear neurons shows a significant reduction upon fear extinction learning, indicating that the phasic cue-responsiveness of fear neurons relates to the high fear state of the animal (Figure 22d). A distinct set of neurons shows the exact opposite pattern of activity: fear extinction neurons do not fire in response to CS^{+}_{av} after the fear conditioning but they show an increase in normalized firing upon presentation of aversive extinguished cues (Figure 22e). As for appetitive extinction neurons, individual fear extinction neurons show a significant increase of their peak firing upon extinction

learning (Figure 22f). Similarly, the specific pattern of activity of fear and fear extinction neurons conveys information about the actual emotional significance of aversively conditioned cues.

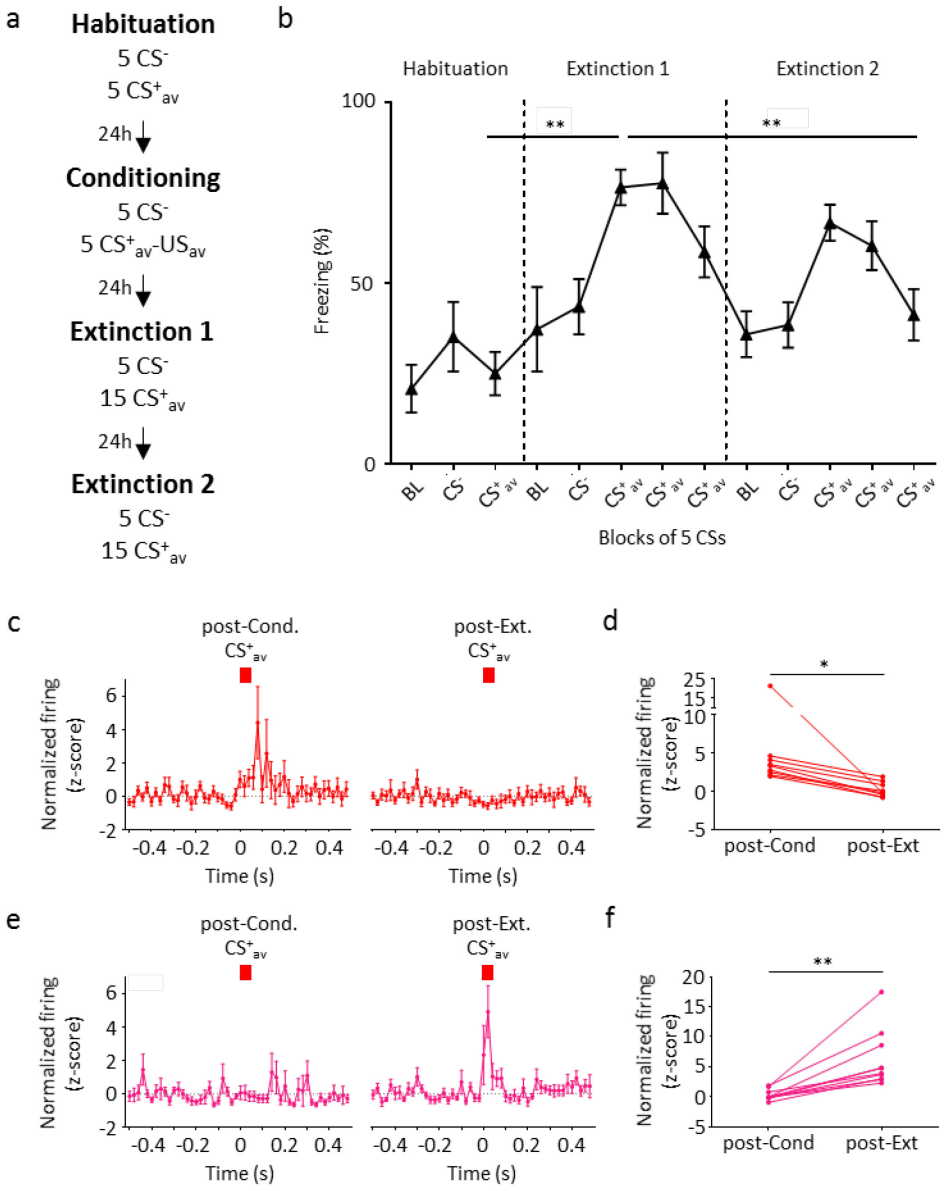

Figure 22. Neuronal correlates of Pavlovian fear conditioning and extinction

a: Behavioral protocol for classical fear conditioning (FC) and extinction. **b**: Freezing levels before Pavlovian fear conditioning (Habituation) and during the two consecutive fear extinction sessions in amygdala-implanted mice (n=6). Before FC, animals show equally low freezing levels to context exposure (BL), CS^- and CS^+_{av} presentations. On the day following FC, mice show increased freezing specifically to CS^+_{av} presentations. Two consecutive days of fear extinction training lead to low freezing levels in response to CS^+_{av}, undistinguishable from pre-conditioning levels. Repeated-measure one-way ANOVA, Tukey correction for multiple comparisons, p<0.01. **c**: Normalized activity (z-score) of fear neurons on high (post-Cond) and low fear states (post-Ext.). Fear neurons (n=9) are specifically CS^+_{av}-excited after FC and not after fear extinction. **d**: Peak of normalized activity for individual fear neurons during high (post-FC) and low (post-FX) fear states. Fear extinction induces a significant decrease of CS^+_{ap}-responsiveness in fear neurons. Two-tailed paired t-test, p<0.05. **e**: Normalized activity (z-score) of fear extinction neurons after Pavlovian FC (post-Cond) and at the end of fear extinction training (post-Ext.). Fear extinction neurons (n=10) are specifically CS^+_{av}-excited by extinguished aversively conditioned cues. **f**: Peak of normalized activity for individual fear extinction neurons during high (post-FC) and low (post-FX) fear states. Fear extinction induces a significant increase in the cue-responsiveness of fear extinction neurons. Two-tailed paired t-test, p<0.01. Error bars indicate mean ± s.e.m.; **: p<0.01

In summary, these results demonstrate that the BA encompasses neurons specifically responding to emotionally relevant cues depending on the emotional state of the animal. In the appetitive domain, appetitive neurons are CS-responsive specifically during high hedonic states whereas appetitive extinction neurons only respond to extinguished appetitive cues. In the aversive domain, similar neuronal types were identified: fear neurons which are cue-responsive during high fear states and fear extinction neurons solely activated by aversively conditioned cues after fear extinction.

Importantly, this data confirms evidence gathered by prior studies showing that the amygdala is recruited by appetitive stimuli as well as aversive ones[71,95] and is important for the encoding of appetitive associative memories. In addition, our work identifies for the first time the population of appetitive extinction neurons which might play an important role in the encoding of appetitive inhibitory learning.

Finally, the single unit recordings of neuronal activity in the BA demonstrate that emotional associative learnings of opposite valence are encoded in a similar fashion: distinct neuronal populations are recruited by conditioned and extinguished cues, both in the aversive and appetitive domain. However, it remains unclear how opposite valences are represented relative to each other. The following section addresses this question of the overlap between neurons recruited by appetitively or aversively conditioned cues.

Relative organization of amygdala circuits of conditioning and extinction of opposite valence

In order to investigate how conditioned and extinction memories for aversive and appetitive valences are represented relative to each other in BA circuits, Pavlovian conditioning and extinction training for both valences were sequentially performed in the same amygdala-implanted animals. Mice were either subjected to appetitive training (conditioning + extinction) followed by aversive training (conditioning + extinction) or to the opposite learning sequence (Figure 23a). Similar to mice which underwent only one conditioning episode of a given valence, mice submitted to sequential appetitive-aversive training acquire valence-specific CR upon both conditioning sessions (Figure 23b). After fear conditioning, CR_{av} levels increase during the presentation of aversive cues and decline through fear extinction training. Likewise, consecutive to appetitive conditioning the level of CR_{ap} increases in response to CS^{+}_{ap} presentations and rapidly declines within one session of appetitive extinction. Similar to single valence training, Mice do not exhibit CRs to the context or to the CS^{-} after conditioning (Figure 23b), indicating that CS^{+}s have gained a specific emotional valence through their association with positively or negatively valenced reinforcers. Importantly, Mice are also able to discriminate between the valence of the two CS^{+}s: no freezing behavior is expressed during the presentations of appetitively conditioned cues and HTR remains low during CS^{+}_{av} presentations (Figure 23b). This strongly demonstrates that there is no cross-talk between these two behaviors, that mice acquired valence-specific memories associated with each specific cue and that they express specific behaviors adapted to the valence of the cues.

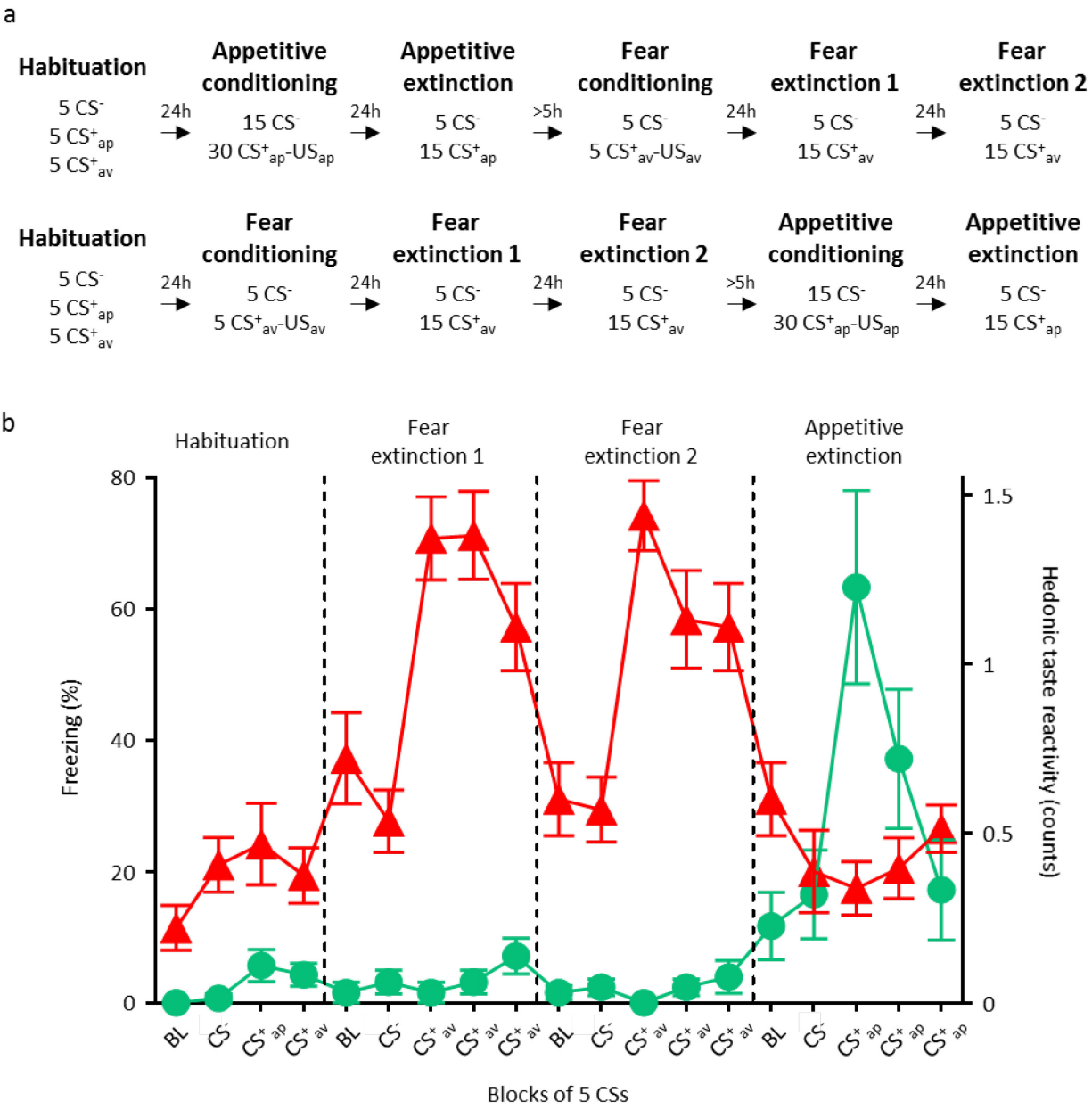

Figure 23. Combined conditioning and extinction of opposite valences

a: Behavioral protocol for combined Pavlovian appetitive and aversive conditioning and extinction. One group of animals was submitted to appetitive learning first (top row) whereas the second group of mice underwent fear learning prior to appetitive conditioning (bottom row). **b**: Appetitive (hedonic taste reactivity, HTR) and aversive (freezing) conditioned responses of animals which underwent double valence conditioning (n=15). HAB, habituation, FX1, first fear extinction session; FX2, second fear extinction session; AX, appetitive extinction. Hedonic taste reactivity (green circles) remains at a very low level until CS^{+}_{ap} presentations after Pavlovian appetitive conditioning. Hedonic taste reactivity decreases very fast in a within-session fashion during appetitive extinction. Likewise, freezing levels are similarly low during habituation on context exposure (BL) and during CS presentations. Fear conditioning induces an increase of freezing levels specific to the CS^{+}_{av} presentations. This high freezing level decreases during fear extinction. Hedonic taste reactivity measured during fear extinction remains very low and freezing level during appetitive extinction is similar to pre-FC levels. Error bars indicate mean ± s.e.m.

Single unit recordings were performed while the animals were submitted to sequential appetitive and aversive conditioning and extinction. The chronicity of the single unit recordings combined with the sequential training across both valences allows for following neurons through the different behavioral sessions and thus makes it possible to address the question of how conditioning and extinction circuits of opposite valences overlap in the BA. As described above, we identified fear neurons, fear extinction neurons, appetitive neurons and appetitive extinction neurons in the BA which activity relates to excitatory and inhibitory learning of appetitive and aversive valences. Table 2 summarizes the total number of neurons in each neuronal subclass.

	Appetitive neurons	Appetitive extinction neurons	Other
Fear neurons	1	2	10
Fear extinction neurons	3	6	16
Other	13	6	111

Table 2: Number of neurons belonging to each class of activity pattern during high and low emotional states of opposite valence. n=168

Testing for the contingency between these neuronal subpopulations reveals a significant association between fear/fear extinction and appetitive/appetitive extinction circuits (Chi-square test, p= 0.02). This indicates that the actual number of neurons in each category differs from a theoretically expected distribution among these neuronal classes. A more detailed analysis of the contingency between these neuronal subpopulations reveals a significant overlap between extinction neurons of opposite valence (Figure 24) (Fisher exact test, p=0.008). No significant overlap is however detected between conditioning neurons of opposite valence or between conditioning neurons of one valence and extinction neurons of the opposite valence (Figure 24). These results suggest that conditioning and extinction of opposite valences mostly rely on distinct neuronal populations in the BA and that only a small fraction of neurons participate to valence-free extinction mechanisms.

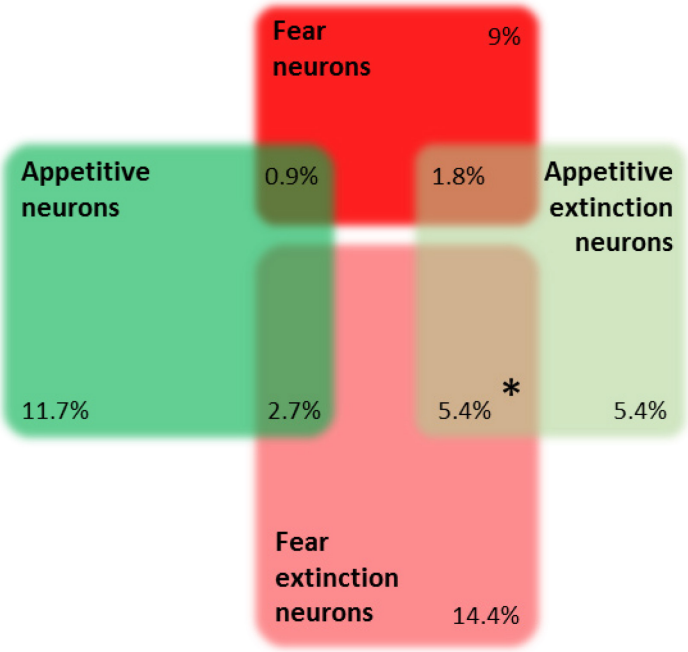

Figure 24. Relative representation of conditioning and extinction of opposite valences in BA circuits

Venn diagram representing the overlap between appetitive (A), appetitive extinction (AX), fear (F) and fear extinction (FX) neurons. Percentages correspond to the proportion of each individual subpopulation over the total of neurons chronically recorded from habituation to the last extinction session (n=168). The asterisk represents the significant association between extinction neurons of opposite valences (Fisher exact test, p=0.008).

Effect of prior appetitive experience on subsequent aversive associative learning episodes

Beyond the question of valence interaction at the neuronal level, numerous studies have suggested that prior experience affects subsequent emotional experience. However, these studies focus mostly on the behavioral effect of the conditioning of one valence on a subsequent conditioning of the opposite valence (i.e. counterconditioning[90–92]) or investigate the effect of US re-evaluation on subsequent performance by manipulating the interoceptive state of the animals (i.e. conditioned taste aversion [142], satiety devaluation [143], appetite revaluation [144]). Currently, no information is available in regard to how neuronal activity at the single cell level is influenced by prior emotional experience. We thus studied the effect of prior appetitive learning episodes on subsequent fear conditioning and extinction, both at the behavioral and at the neuronal level.

After the habituation session, one group of animals was submitted to appetitive conditioning and extinction followed by fear conditioning and fear extinction whereas the second group of mice underwent fear conditioning and extinction without being exposed to prior appetitive experience (Figure 25a). To assess that mice submitted to the sequential conditioning of both valences actually acquired appetitive learning, their CR_{ap} was analyzed during the appetitive extinction session. As described in previous paragraphs, mice show high levels of HTR on CS^{+}_{ap} after appetitive conditioning and this behavioral response declines during extinction training (Figure 25b). We then looked at aversive CR during the fear conditioning phase to determine whether, as for

counterconditioning paradigms, prior appetitive conditioning would delay the acquisition of aversive CRs.

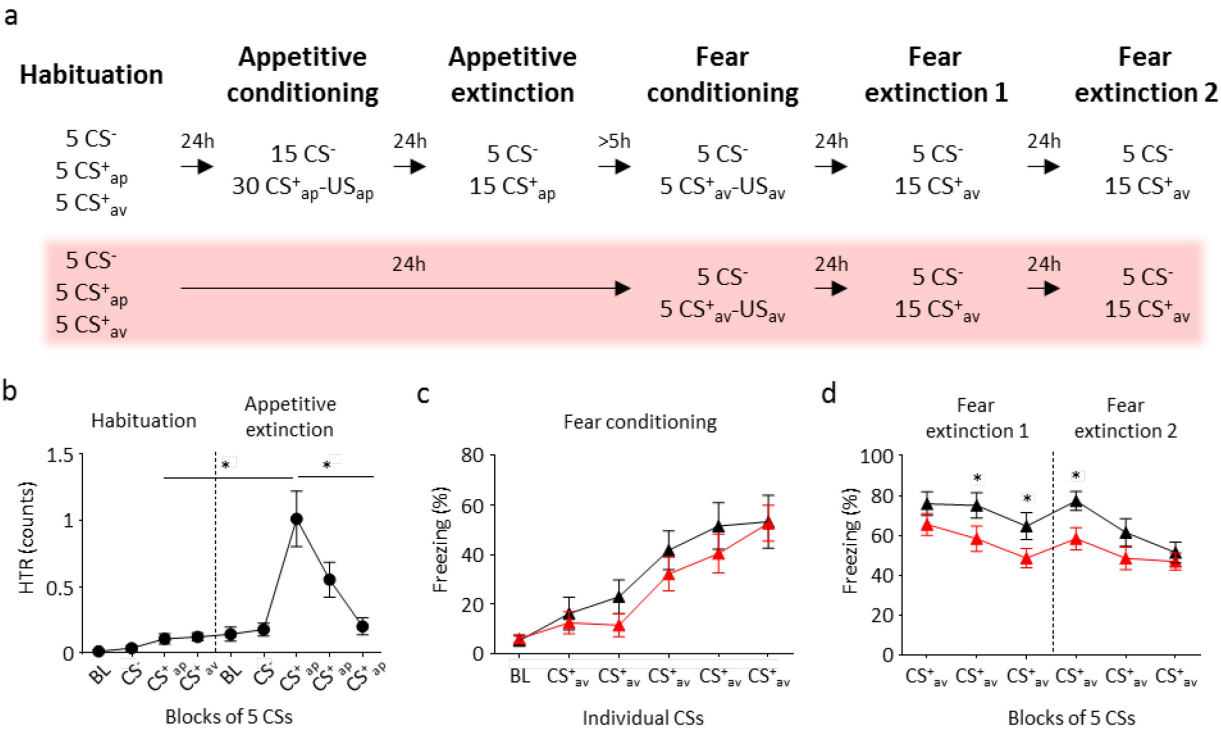

Figure 25. Effect of prior appetitive experience on aversive learning

a: Behavioral protocol. One group of mice was submitted to appetitive conditioning and extinction before undergoing fear conditioning and fear extinction (n=17). The second group of animals was only subjected to fear conditioning and fear extinction (n=23). **b**: Appetitive learning in the group of animals submitted to double valence conditioning. Before Pavlovian appetitive conditioning (HAB, habituation), mice do not show hedonic taste reactivity. After Pavlovian appetitive conditioning (AX, appetitive extinction), mice show a significant increase in HTR to the appetitively conditioned cue (CS^{+}_{ap}), but not to the context (BL, baseline) or to the non-reinforced cue (CS^{-}). Appetitive extinction leads to a progressive decrement in HTR in a within-session manner. Repeated-measure one-way ANOVA, Tukey correction for multiple comparisons, $p<0.0001$. **c**: Freezing behavior elicited by the context (BL, baseline) and the individual CS^{+}_{av} during the fear conditioning session for animals which underwent prior appetitive experience (black triangles, n=17) and animals which were only undergoing fear learning (red triangles, n=23). Mice from both groups acquire fear conditioning at a similar speed. Two-way ANOVA, Sidak correction for multiple comparisons, $p=0.8$. **d**: Freezing behavior evoked by CS^{+}_{av} presentations during the two fear extinction sessions (FX1, FX2) for the fear only group (in red) and the group of animals previously exposed to appetitive experience (in black). Both groups show the same post-conditioning freezing levels but mice which underwent prior appetitive experience exhibit a delay in the fear extinction acquisition during FX1 and a lack of fear extinction consolidation (first block of FX2). Two-way ANOVA, uncorrected Fisher's LSD test, $p=0.5$. Error bars indicate mean ± s.e.m. *: $p<0.05$.

However, contrary to observations made in counterconditioning procedures, mice submitted to prior appetitive experience have a similar learning curve for fear conditioning than animals which were not exposed to appetitive learning (Figure 25c). This important difference can be explained by the fact that in our experiments, two different CSs were used for appetitively and aversively conditioned dues and by the fact that an appetitive extinction session is interleaved between appetitive and fear conditionings. However, although no difference can be observed between the

two groups during the fear conditioning session, extinction learning is affected by the prior appetitive experience. The two groups of animals show similar levels of post-conditioning freezing but mice which underwent appetitive experience before fear conditioning maintain higher fear responses during the first extinction session and have higher freezing levels on the first CS^{+}_{av} presentations of the second extinction session (Figure 25d). Nevertheless, at the end of the fear extinction training, both groups of animals show similar levels of freezing. These results suggest that prior appetitive experience interferes with fear extinction learning and consolidation (Figure 25d). We then investigated the neuronal correlates of this delay in fear extinction induced by prior appetitive experience. Two groups of mice were implanted with single unit electrodes in the BA. One group was submitted to combined appetitive and aversive learning while the other group only underwent fear conditioning and extinction. Behavioral procedures were identical to the ones previously described in this paragraph. Fear neurons and fear extinction neurons were identified in both behavioral groups. Comparison of the activity of fear cells in response to CS^{+}_{av} during high fear state does not differ between groups (Figure 26a). In contrast, the activity of fear extinction neurons in the group of animals which received prior appetitive experience shows a significant reduction compared to the fear only group (Figure 26b). The proportions of these two neuronal populations seem to be unaffected by prior experience, indicating that prior appetitive experience does not lead to the recruitment of a smaller pool of neurons into fear and fear extinction memory traces but rather modulates the activity level of cells involved in the encoding of these memories (Figure 26c).

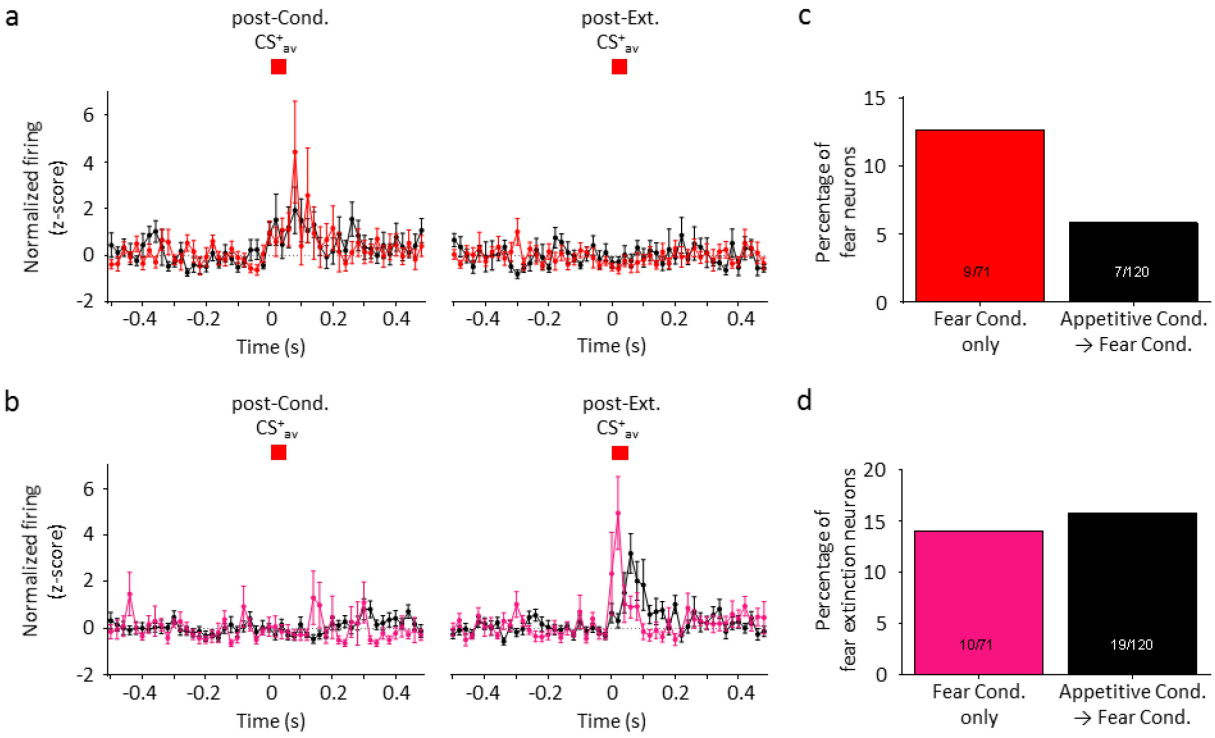

Figure 26. Effect of prior appetitive experience on aversive circuits

a: Normalized activity (z-score) of fear neurons during high (post-Cond) and low (post-Ext.) fear states in the fear only group (in red, n=9) or in the group of mice which received prior appetitive experience (in black, n=7). Post-conditioning CS^{+}_{av}-responsiveness does not differ between the two behavioral treatments. Two-way ANOVA, Sidak correction for multiple comparisons, p=0.8. **b**: Normalized activity of fear extinction neurons after FC (post-Cond) and fear extinction (post-Ext.) for group of animals which only received FC training (in pink, n=10) or for mice which underwent prior appetitive experience (in black, n=19). The normalized firing in response to aversively conditioned cues of fear extinction neurons during low fear states in mice which had prior appetitive training is significantly lower than the activity for the fear conditioning only group. Two-way ANOVA, Sidak correction for multiple comparisons, p<0.0001. Error bars indicate mean ± s.e.m. **c**: Percentages of fear (left panel) and fear extinction neurons (right panel) over the total population recorded in the BA during habituation and the two fear extinction sessions. Proportions of fear neurons do not differ between the fear only group (FC, in red) and the group which received prior appetitive experience (AC+FC, in black). Two-tailed Z-test, p=0.1. Similarly, no difference can be found between the proportion of fear extinction neurons from the fear only group (FC, in pink) and from the prior appetitive learning group (AC+FC, in black). Two-tailed Z-test, p=0.7.

Consistent with the behavioral effects described above, prior appetitive experience specifically influences fear extinction encoding. Taken together this data suggests that prior appetitive conditioning and/or prior appetitive extinction lead to a reduction in the aversive cue-responsiveness of fear extinction neurons translating at the behavioral level by a delay in fear extinction learning.

Asymmetric recruitment of amygdala circuits by aversively and appetitively conditioned cues

Remarkably, at the behavioral procedures leading to appetitive and aversive learning are asymmetric. The acquisition of appetitive conditioning requires 30 pairings between the CS^{+}_{ap} and the US_{ap}, whereas fear conditioning is acquired after only 5 CS^{+}_{av}- US_{av} pairings. In contrast, appetitive extinction occurs within one session while fear extinction requires two sessions of 15 non-reinforced CS^{+} presentations (Figure 27a). The faster acquisition of the aversive conditioning and the slower acquisition of aversive extinction could be due to the difference in the biological relevance of the USs. Indeed, from an ethological point of view, avoiding threats seems more crucial to survival than ceasing food resources.

In order to investigate the BA neuronal correlates of this asymmetry, we compared the neuronal activity evoked by appetitively conditioned cues to aversive ones. We found the CS-evoked activity in the BA to be strongly biased towards aversive valence. At the population level, the normalized cue-responsiveness of CS^{+}-excited neurons shows a strong asymmetry between CS^{+}_{ap} and CS^{+}_{av} (Figure 27b). The phasic excitation elicited by the cue previously paired with the foot-shock is approximately two times higher than the one evoked by the cue previously associated with the sucrose delivery (Figure 27c, left panel). Interestingly, this aversive bias is not sensitive to extinction training, post-extinction cue-responsiveness to CS^{+}_{ap} being approximately half of CS^{+}_{av}-induced excitation (Figure 27c, right panel). In addition to the aversive bias of the BA CS^{+}-evoked activity, the proportion of neurons recruited by aversively conditioned cues strongly differs from the one for appetitively conditioned cues: the proportion of CS^{+}_{av}-excited neurons is almost two times larger than the one of the CS^{+}_{ap}-excited neurons (Figure 27d).

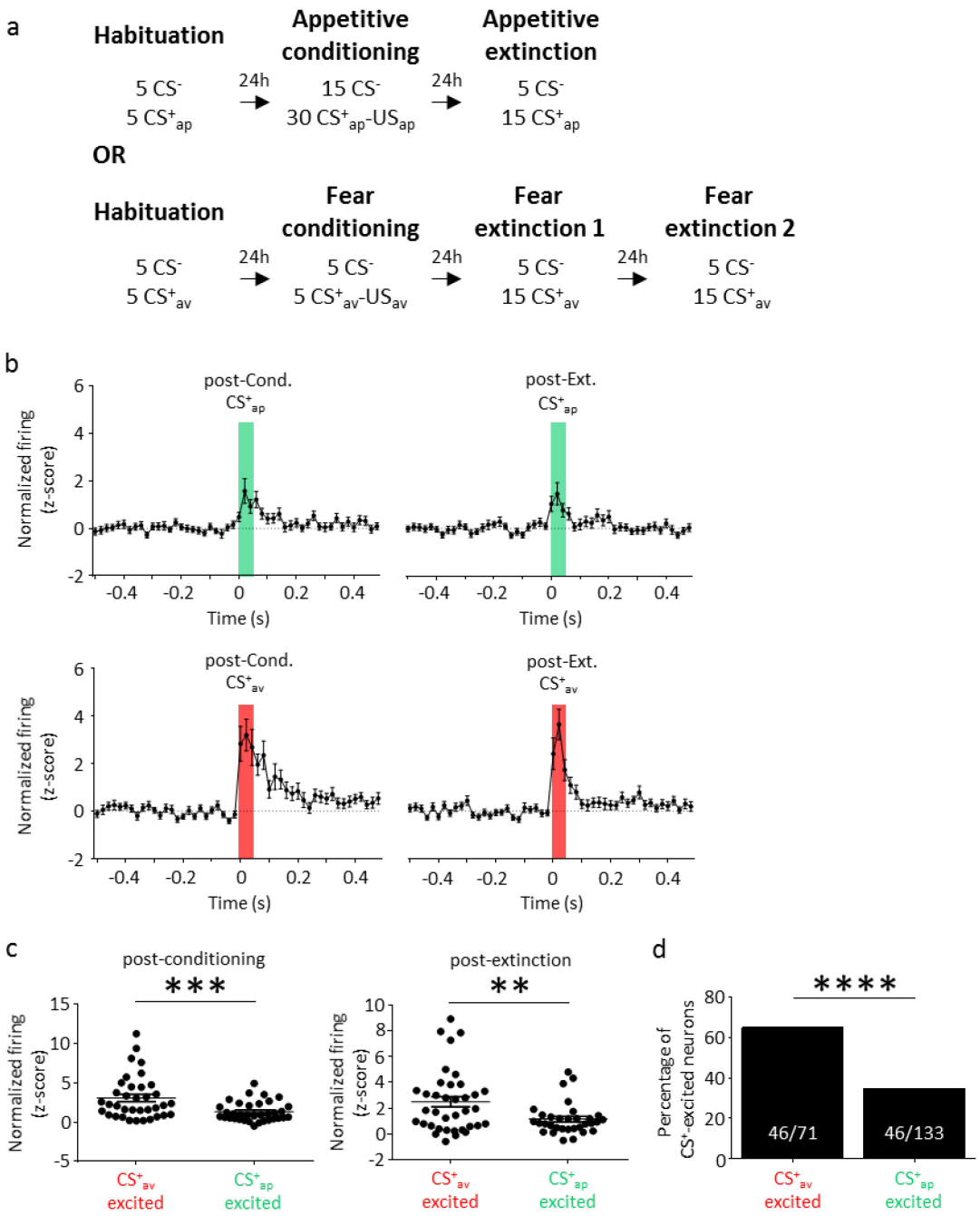

Figure 27. Asymmetric recruitment of BA circuits by opposite emotional valences

a: Behavioral protocol. Mice were either submitted to appetitive conditioning and extinction (n=11) or to fear conditioning followed by fear extinction (n=6). **b**: Averaged normalized activity in BA in response to CS^{+}_{ap} (top panels, n=46) and CS^{+}_{av} (bottom panels, n=46) post-conditioning (post-Cond) and post- extinction (post-Ext.). Only CS^{+}-excited cells were included. Neuronal responsiveness to CS^{+}_{av} is much more elevated than that to CS^{+}_{ap} both after conditioning and extinction of opposite valences. **c**: Normalized activity of individual CS^{+}_{av}- and CS^{+}_{ap}-excited neurons after conditioning and extinction averaged over 100 ms after peep onset. Both post-conditioning and post-extinction activity evoked by CS^{+}_{av} is significantly larger than the one induced by CS^{+}_{ap} (Two-tailed unpaired t-test, p=0.0009 for post-conditioning and p=0.006 for post-extinction). **d**: Percentages of neurons recruited by aversively or appetitively conditioned cues over the total of neurons chronically recorded during single valence learning. The proportion of neurons recruited by CS^{+}_{av} is significantly larger than the one recruited by CS^{+}_{ap} (Two-tailed Z-test, p<0.0001).

In addition to conditioning and extinction neurons of opposite valences, the BA also contains another class of neurons which cue-responsiveness is insensitive to extinction training. This class of neurons exhibits a CS^+-induced excitation after conditioning and maintains a significant cue-responsiveness to extinguished cues. This pattern of activity is found in both valences: appetitive extinction-resistant neurons being excited by appetitively conditioned cues both after appetitive conditioning and appetitive extinction (Figure 28a) and fear extinction-resistant neurons being CS^+_{av}-responsive both after fear conditioning and fear extinction (Figure 28b).

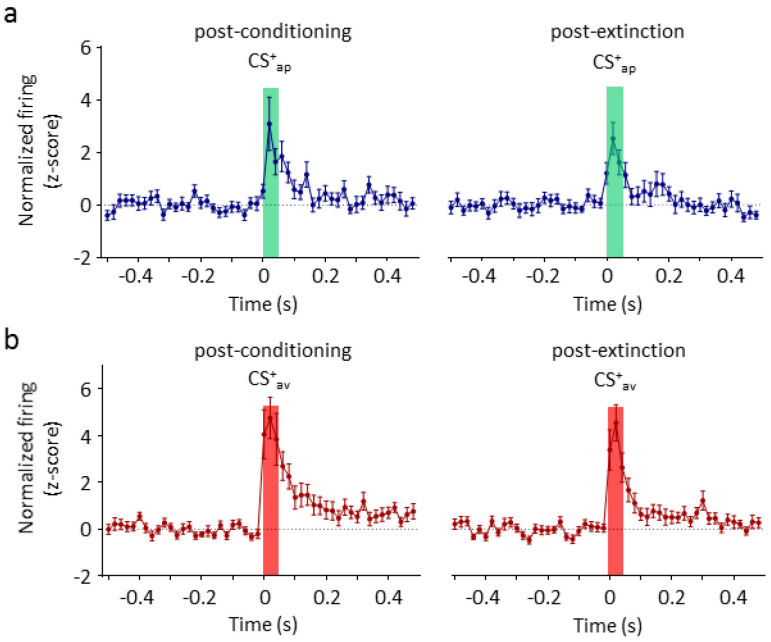

Figure 28. Extinction resistant neurons of opposite valence
Normalized activity (z-score) of extinction-resistant neurons, which exhibit a specific pattern of activity excited by CS^+ post-conditioning (post-Cond), maintained to extinguished cues (post-Ext.) **a**: Appetitive extinction neurons (n=21). **b**: Fear extinction neurons (n=27). Error bars indicate mean ± s.e.m.

In order to determine which particular neuronal subpopulations are responsible for such a strong aversive bias in the BA activity, we compared between both valences the proportions and peak activity of conditioning, extinction, and extinction-resistant neurons. No difference in proportion can be found between neurons recruited by CS^+s exclusively after conditioning (fear versus appetitive neurons; Figure 29a, left panel). Similarly, the proportions of fear extinction and appetitive extinction neurons do not differ from each other (Figure 29a, middle panel). However, the proportion of fear extinction-resistant neurons is more than twice larger than the proportion of appetitive extinction-resistant neurons (Figure 29a, right panel). Consistent with these observations, the level of CS^+-evoked activity is not different between conditioning neurons of opposite valence (Figure 29b, left panel) or for extinction neurons (Figure 29b, middle panel). However, the CS^+-responsiveness of fear extinction neurons drastically differs from that of

appetitive extinction neurons: CS^+-induced excitation is, on average, twice larger for fear extinction-resistant neurons than for appetitive extinction neurons (Figure 29b, right panels).

Taken together these results indicate that the strong aversive bias in amygdala activity relies on extinction-resistant neurons, a neuronal population which cue-responsiveness is insensitive to extinction training.

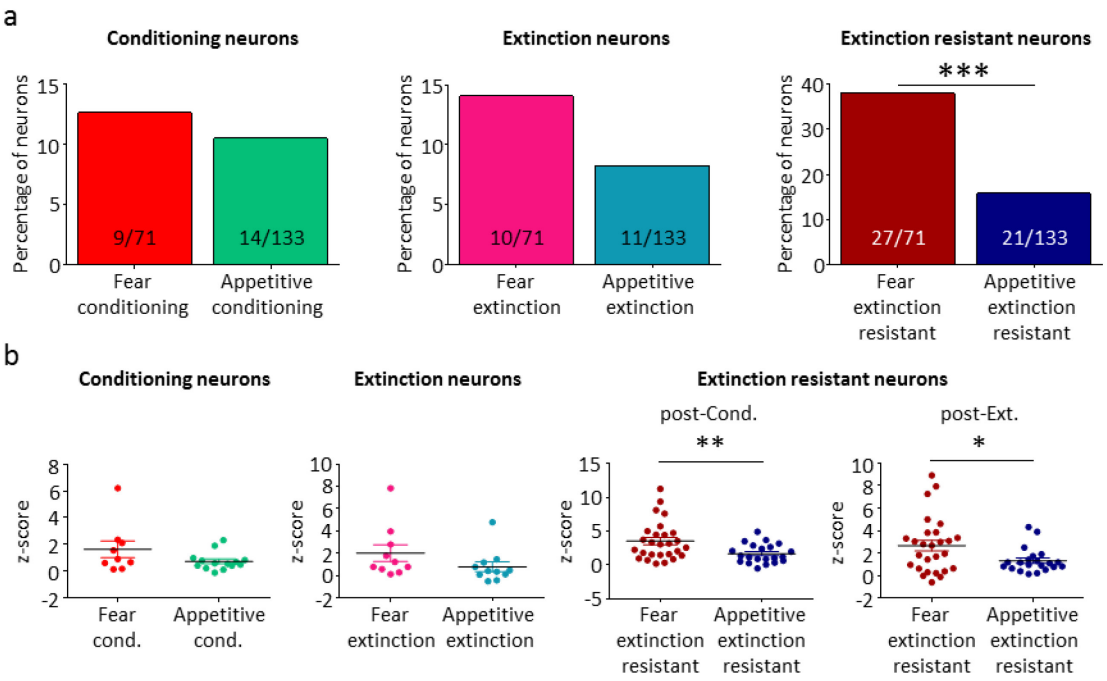

Figure 29. Extinction resistant neurons are responsible for the BA aversive bias

a: Percentage of conditioning neurons (F, fear and A, appetitive; left panel), extinction neurons (FX, fear extinction and AX, appetitive extinction; middle panel) and extinction resistant neurons (FXR, fear extinction resistant and AXR, appetitive extinction resistant; right panel) over the total of chronically recorded neurons during single valence learning. Only the proportion of extinction resistant neurons of opposite valences shows a significant difference, more FXR being recruited by CS^+_{av} than AXR by CS^+_{ap} (Two-tailed Z-test, $p<0.001$). **b**: Normalized activity of individual conditioning neurons (F, fear and A, appetitive; top left panel), extinction neurons (FX, fear extinction and AX, appetitive extinction; top right panel) and extinction resistant neurons (FXR, fear extinction resistant and AXR, appetitive extinction resistant; bottom panels) averaged over 100 ms after peep onset. Conditioning and extinction neurons of opposite valence show similar level of cue-responsiveness, contrary to extinction resistant neurons for which fear extinction resistant neurons CS^+-evoked activity is significantly higher than the one of appetitive extinction neurons. (Two-tailed unpaired t-test, $p=0.009$ for post-conditioning and $p=0.03$ for post-extinction).

We hypothesized that if the BA aversive bias was linked to the asymmetrical biological relevance of the USs then the neurons responsible for this bias should be valence-specific, i.e. fear neurons should only be responsive to aversively conditioned cues while appetitive extinction neurons should only show excitation upon the presentation of CS^{+}_{ap}. In order to address this question, single unit recordings were performed in amygdala-implanted mice which underwent combined aversive and appetitive learning (Figure 23a, b) and the activity of extinction-resistant neurons was followed during the different learning phases. Table 3 summarizes the overlap and segregation of extinction-resistant neurons of opposite valences.

	Appetitive extinction resistant neurons	Other
Fear extinction resistant neurons	9	32
Other	13	114

Table 3: Overlap between extinction resistant neurons of opposite valences (total number of neurons). n=168

Despite the existence of a small overlap between appetitive and fear extinction-resistant neurons (Figure 30), no significant association can be detected between extinction-resistant neurons of opposite valence (Fisher exact test, p= 0.0645).

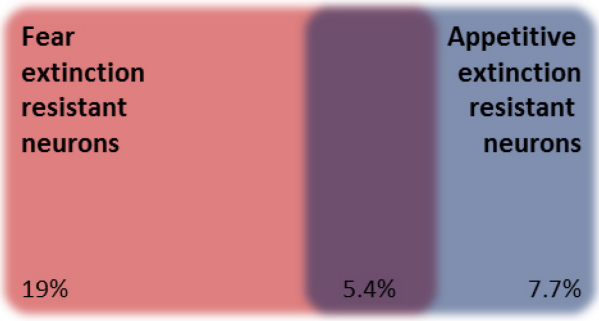

Figure 30. Extinction resistant neurons are mostly segregated
Venn diagram representing the overlap between appetitive (AXR) and fear extinction resistant neurons (FXR). Percentages correspond to the proportion of each subpopulation over the total of chronically recorded neurons during combined valence learning paradigms (n=168). No significant association can be detected between extinction resistant neurons of opposite valence (Fisher exact test, p= 0.0645).

These results suggest that extinction-resistant neurons might participate in the maintenance of conditioning memories traces over extinction learning in a valence-specific manner and are extremely sensitive to the salience of the USs.

Discussion

Pavlovian appetitive conditioning

The accurate comparison of the neuronal circuits recruited by opposite valences requires using behavioral paradigms relying on the same type of association between CSs and USs for both valences. In order to achieve this goal, I developed an appetitive conditioning in mice based on a purely Pavlovian basis. In this paradigm, as for Pavlovian fear conditioning, animals receive the US upon the CSs presentation, independently of their behavior, i.e. with no requirement to perform any action for the US to be delivered. This was achieved by the development of surgical procedures for the implantation of intra-oral cannulae which allow the delivery of palatable solutions directly into the oral cavity. In addition, the accurate evaluation of associative emotional learning crucially relies on the appropriate choice of the behavioral responses used to assess the emotional states of an animal. For the first time, we characterized appetitive conditioned responses in mice acquired on a purely Pavlovian basis. We demonstrated in the present study that a specific type of orofacial movements, called hedonic taste reactivity (HTR), represents the actual emotional significance gained by the CS through its pairing with the appetitive US. HTR are indeed expressed both as appetitive URs and CRs, they are specific of the valence of the environmental cues (they are exclusively expressed during the presentation of the CS^{+}_{ap} and not during CS^{-} nor context exposure and they are sensitive to extinction training) and correlate with the relative emotional valence of the CS, as shown by their sensitivity to the US_{ap} intensity. The development of this Pavlovian appetitive conditioning has been the ground on which we were then able to accurately study the relative representation of opposite valences in amygdala circuits.

Representation of opposite valences in amygdala circuits

Inactivation and lesions studies have demonstrated that the BLA participates to the attribution of an emotional significance to otherwise neutral cues through their contingent occurrence with biologically relevant events and have thus shown that the BLA is involved in emotional associative learning of both positive and negative valence. More recently, few electrophysiological studies have indicated that the amygdala comprises single neurons responding to either aversive or appetitive cue, suggesting a valence encoding in distinct neuronal circuits in the BLA[71]. However, the representation of excitatory and inhibitory learning of opposite valences in the amygdala has not

yet been investigated. We were able to finally address this question by combining classical fear conditioning and our newly developed classical appetitive conditioning. The present study pinpoints the neuronal circuits involved in excitatory and inhibitory learnings of opposite valences. We identified fear neurons and fear extinction neurons specifically responding to aversive cues after conditioning or after extinction. Interestingly, we found appetitive learning to be represented in a similar manner in BA circuits: appetitive neurons being CS_{ap}-responsive on high hedonic states and appetitive extinction neurons being excited by extinguished CS_{ap}. Prominently, this study is the first evidence of the existence of appetitive extinction neurons characterized on a purely Pavlovian basis.

The identification of these different neuronal populations has allowed us to address the question of the relative representation of excitatory and inhibitory learning of opposite valences in amygdala circuits. Similar to a previous study[71], we found valence to be generally represented in a segregated fashion in the BA, appetitive neurons being mainly non-overlapping with fear neurons. Likewise, extinction of opposite valences mostly recruits distinct neuronal subpopulations. Our results thus suggest that amygdala encoding is not only specific for excitatory and inhibitory learning but also for the valence of these different learning types.

However, we have also identified a small subpopulation of valence-free extinction neurons, indicating that the processes involved in extinction learning of opposite valences could at least partially have common neuronal substrates. The existence of such a neuronal population suggests that inhibitory learning might, in contrast to excitatory learning, rely on a synergy between valence-dependent and valence-specific circuits. Specific manipulations of valence-free and valence-dependent extinction neurons would need to be performed to understand the relative participation of these two populations in extinction learning.

Remarkably, we did not detect a significant association between conditioning neurons of one valence and extinction neurons of the opposite valence. This demonstrates that at the single cell level, extinction of one valence is not similar to the conditioning of the opposite, suggesting that appetitive extinction is not aversive and that aversive extinction is not rewarding *per se*.

The segregation of these neuronal subpopulations raises the question of whether this is due to the fact that USs of opposite valences used in this study were of different sensory modalities. To control for this, we have tried to develop a Pavlovian aversive conditioning on a gustatory modality, using

intra-oral delivery of quinine or acetic acid as US_{av}. Although mice do show aversive taste reactivity when exposed to these tastants, we were not able to assess gustatory aversive conditioning as no taste reactivity could be observed during the memory retrieval tests (data not shown). Nevertheless, although some modality-specific neurons were identified in the BLA, this structure is well-known to contain multimodal neurons on which relies the actual function of the BLA in linking environmental cues to emotionally relevant events[145]. Additionally, it was shown more recently that similar valences involving different sensory modalities are more likely to recruit the same BLA neurons than opposite valences of the same sensory modality[146].

The present study is mainly correlative and thus to definitely show a causal link between the activity of these discrete neuronal populations and conditioning and extinction of opposite valences, specific manipulations of their activity remain necessary. So far no specific molecular marker has been identified to characterize these different neuronal populations. Nevertheless, the participation to different long-range circuits of these neurons could be an entry point for specific manipulations based on optogenetic approaches. A very recent publication has used this strategy and showed that optogenetic stimulation of BLA neurons projecting to the nucleus accumbens was sufficient to induce instrumental appetitive behavior[147]. This would suggest that appetitive neurons identified in our study could project to the nucleus accumbens. Additionally, work from our lab described in the first part of the results section[63] has shown fear neurons to be preferentially projecting to the prelimbic division of the medial prefrontal cortex whereas fear extinction neurons send axons to the infralimbic division. Therefore, optogenetic experiments taking advantage of the distinct long-range connectivity of these discrete neuronal populations could allow to causally link their activity to appetitive and aversive behaviors.

In summary, conditioning and extinction are mostly encoded in a valence-specific manner in the BA circuits, confirming the role of this structure not only in the general process of linking an environmental cue to biologically relevant event but actually supporting learning about the current specific valence of stimuli in an ever changing environment. In addition, valence-specific neurons for excitatory and inhibitory learning might be part of distinct long-range circuitry allowing for specific behavioral adaptation upon changes in environmental circumstances.

Emotional learning in context

Behavioral studies have shown that the expression of behavioral responses highly depends on the context in a general sense, i.e. not only the physical context (the arena/surroundings in which emotional experience takes place) but on its interaction with the internal state of an animal[141,143,144,148–153]. This suggests that prior experience influences how emotionally relevant events are perceived and memorized. However, so far, no study investigated how neuronal circuits would process emotionally relevant stimuli depending on the prior experience of animals. This study is the first to describe the modulation by prior emotional experience of neuronal circuits implicated in emotional learning. Here, we show that contrary to what was observed in counterconditioning paradigms, prior appetitive experience does not lead to a delay in the acquisition of a subsequent fear conditioning episode. This difference might be due to two different factors. In counterconditioning, the same cue is sequentially associated with the USs of opposite valences. Animals have therefore to learn at the same time that the CS does not predict a reward and that the CS is predictive of a foot-shock delivery. By using two different CSs for appetitive and aversive conditioning and by extinguishing the appetitive memory before fear conditioning, we prevented the valence competition for the CS significance which occurs in counterconditioning. Consequently, in our experiments, animals submitted to prior appetitive experience acquire FC as fast as animals which only underwent fear conditioning. This result is consistent with our observation that conditioning circuits of opposite valence and conditioning circuits of one valence and extinction circuits of the opposite are mostly segregated and with the fact that fear neurons activity is not modulated by prior appetitive experience.

In contrast to the absence of modulation of FC acquisition by prior appetitive learning, fear extinction learning was affected by prior emotional episodes. Compared to animals only exposed to fear conditioning, mice which underwent prior appetitive conditioning and extinction show a delay in fear extinction learning on the first day of extinction training and a lack of fear extinction consolidation on the following day. Consistent with these behavioral observations, we found fear extinction neurons to have a reduced cue-responsiveness in mice which underwent appetitive learning first compared to animals which only received FC training.

In summary, prior appetitive experience modulates specifically fear extinction both at the behavioral and at the neuronal level, reducing behavioral fear extinction and fear extinction

neurons activity. The participation of long-range circuits in the effect of prior experience still remains to be investigated. In particular differential modulations of projections to the prelimbic and infralimbic division of the prefrontal cortex by prior appetitive experience might contribute to the effect we observed on fear extinction consolidation. In addition, considering the predominant role of the hippocampus in autobiographical memories and in particular of the ventral hippocampus in providing contextual information to BLA circuits, investigating the interaction between this structure and the amygdala in the framework of the influence of prior emotional episodes would be critical to our understanding of the long-range circuit mechanisms involved in emotional hysteresis.

Aversive bias in amygdala circuits

Consistent with the dominance of aversive paradigms in the study of amygdala circuits, we found the overall activity of the BA to be strongly biased towards aversive stimuli. Indeed, CS-evoked neuronal excitation is twice larger for cues which were paired with footshocks than for appetitively conditioned cues. Two factors are responsible for this aversive bias. First CS^{+}_{av} recruit a larger proportion of BA neurons than CS^{+}_{ap}. Second, neurons recruited by the aversively conditioned cues have a higher level of cue-responsiveness than CS^{+}_{ap} excited neurons. Remarkably, this difference of activity between CS^{+}_{ap} and CS^{+}_{av}-excited neurons is maintained post-extinction. Detailed analysis of the neuronal subpopulations contained in the BA revealed that conditioning neurons (appetitive and fear neurons) and extinction neurons (fear extinction and appetitive neurons) are not involved in the BA aversive bias. However, a third class of neurons was also identified in the BA. These neurons, called extinction-resistant neurons, are CS^{+}-excited during both post-conditioning and post-extinction periods. The aversive bias of this neuronal population suggests that they are responsible for the overall BA aversive bias. Indeed, both the cue-responsiveness and the proportion of extinction-resistant neurons show a strong asymmetry in response to appetitively and aversively conditioned cues in favor of aversiveness.

Fear extinction-resistant neurons have already been identified by us and others[61–63] and they have been hypothesized to be involved in the maintenance after the extinction of the fear memory trace. An alternative explanation for the persistent activity of extinction-resistant neurons after extinction training would be that these neurons act as salience-detector without carrying any information

relative to mnesic processes or to valence. The specific involvement of extinction-resistant neurons in BA aversive bias could support the "salience encoding" hypothesis. Indeed from an ethological point of view, it is much more crucial to the survival of organisms to avoid threats than to cease opportunities for food resources. Thus, we can infer that the salience of an aversive event such as a footshock is much larger than the one of an intra-oral delivery of a sucrose solution. Consistent with this view, we found Pavlovian appetitive conditioning to require six times more CS-US pairings to be acquired than fear conditioning.

In order to discrepate between these two hypotheses, we tested the valence-specificity of extinction-resistant neurons, postulating that if these neurons were involved in salience detection rather than in emotionally-valenced memory traces, extinction-resistant neurons of opposite valence would be overlapping. Our chronic single unit recordings during combined valence learning show that extinction-resistant neurons of opposite valences only partially overlap and are mostly segregated. The lack of significant association between these neuronal populations speaks in favor of a preferential role of extinction-resistant neurons in the conservation of valence-specific memory traces after extinction. Consistent with this interpretation, we found at the behavioral level a strong asymmetry between the extinction of opposite valences, fear extinction requiring two-time more training than appetitive extinction. Accordingly, we also found aversive spontaneous recovery to be much more important than appetitive one (data not shown).

Materials and methods

Animals

Male C57BL6/J mice (2 to 4 months old; Harlan Ltd.) were individually housed 7 days before any experimental procedure under a 12h light/dark cycle. Mice were provided with food *ad libitum* during the entire experiment. All experiments were performed during the light cycle. All animal procedures were performed in accordance with institutional guidelines and were approved by the Veterinary Department of the Canton of Basel-Stadt.

Surgical procedures

For all surgeries, mice were anesthetized with isoflurane (induction 5%, maintenance 1.5%; Attane, Provet) in oxygen-enriched air (Oxymat 3, Weinmann). Analgesia was provided by a combination of local injections of ropivacaine (15 µg/g, subcutaneous, Naropin, AstraZeneca) and systemic injections of meloxicam (8 µg/g, intraperitoneal, Metacam, Boehringer). Mice were secured on a stereotaxic frame (David Kopf Instruments, Bilaney GmbH) and body temperature was maintained constant at 36°C by mean of a feedback-controlled heating pad (CMA/150, CMA/Microdialysis).

Intra-oral cannula implantation

To perform appetitive conditioning on a purely Pavlovian manner, mice were implanted unilaterally with intra-oral cannula consisting of a polyethylene tubing (0.58mm inner diameter; Portex Ltd) attached to a cannula (PlasticOne). Intra-oral cannulae were inserted lateral to the first molar, along the zygomatic arch and ended on the skull where they were secured with cyanoacrylate adhesive gel. Intra-oral cannula did not interfere with the normal feeding behavior of the animals and allowed for passive delivery of fluid directly into the oral cavity. After the implantation, intra-oral cannulae were daily flushed with drinking water to prevent clogging and mice were given 7 days of recovery before any other subsequent manipulation.

Electrode implantation

Mice were unilaterally implanted in the basal nucleus of the amygdala with custom-made electrodes consisting of 16 individually insulated, gold-plated nichrome wires (13 µm inner diameter, impedance 50-150 kΩ; California Fine Wire) contained in a 26 gauge stainless steel guide cannula and attached to an 18 pin connector (Omnetics). The electrode was aimed at the following coordinates: 1.6 mm posterior to bregma, ±3.35 mm lateral to the midline, 4.2 mm deep from the cortical surface and secured to the skull with cyanoacrylate adhesive gel. After the electrode implantation, mice were given at least 7 days of recovery before any subsequent manipulation. At the conclusion of the experiment, recording sites were marked with electrolytic lesions made under deep anesthesia and electrode locations were reconstructed with standard histological techniques.

Behavioral procedures

Water restriction

After a 7 days recovery period from surgery, mice were submitted to water restriction in order to unmask behavioral appetitive conditioned responses. Animals received access to water for 30 minutes per day, at the same time of the day in order to ensure similar interoceptive state across the different behavioral sessions. Water restriction was initiated one week before the first behavioral session and maintained until the end of the behavioral training.

Handling

To habituate them to the connection of the head implants for the sucrose infusion and single unit recordings, mice were daily handled for ten to fifteen minutes by the experimenter during the week preceding the first behavioral session.

Contexts

To prevent contextual interferences between memory formation and retrieval of opposite valences, behavioral sessions were taking place in 3 different contexts. Habituation and extinction sessions took place in context A which consisted of a blue sound-attenuated chamber containing a large transparent circular Plexiglas arena with a superelevated transparent Plexiglas floor. Olfactory contextualization was provided by cleaning the context A with acetic acid (1%). Illumination was provided by white light sources located at the bottom of the context. Fear conditioning took place in context B which consisted of a dark gray sound-attenuated chamber containing a transparent square Plexiglas chamber equipped with a grid floor. Context B was cleaned with ethanol (70%) and illuminated with dim white lights located on the top of the chamber. Appetitive conditioning took place in context C consisting of a dark gray chamber containing a small opaque circular Plexiglas chamber with a superelevated transparent Plexiglas floor. Context C was cleaned with water and illuminated with dim light located underneath the platform.

Behavioral protocols

To ensure discrimination between the different auditory cues, sounds of different frequencies were used as CSs associated with foot-shock, with reward and as non-reinforced CSs (3 kHz, 7.5 kHz, 12 kHz or white noise, CS frequencies being randomized across behavioral groups).

On day 1, mice were submitted to a habituation session in context A, in which they received 5 presentations of CS-, 5 presentations of CS^{+}_{ap} and 5 presentations CS^{+}_{av} (each CS consisting of 50 ms pips repeated at 0.9 Hz, total CS duration: 10 s, sound-pressure level: 75 dB). 24h after the habituation session mice were subjected to the first conditioning session, consisting in either appetitive conditioning or fear conditioning.

Appetitive conditioning consisted of 30 paired presentations of CS^{+}_{ap} and US_{ap} and 15 unreinforced presentations of CS^{-}. US_{ap} consisted of an intra-oral delivery of a sucrose solution (volume: 20 µL, concentration: 0.8 or 1 M; Fluka, rate: 5.28 mL per sec, delivered with a Hamilton pump), the onset of the US_{ap} coinciding with the offset of the CS^{+}_{ap}. On the day following appetitive conditioning, mice were submitted to a single appetitive extinction session in context A during which they received 5 CS^{-} presentations followed by 15 un-reinforced CS^{+}_{ap} presentations.

Fear conditioning consisted of 5 paired presentations of CS^{+}_{av} and US_{av} and 5 unreinforced presentations of CS^{-}. A mild footshock was used as US_{av} (intensity: 0.65 mA, duration: 1 s), the onset of the US_{av} coinciding with the offset of the CS^{+}_{av}. On the two consecutive days following fear conditioning, mice were submitted to fear extinction sessions in context A during which they received 5 CS^{-} presentations followed by 15 un-reinforced CS^{+}_{av} presentations.

Behavioral measurements and analysis

Context A and B were equipped with an infra-red beam frame placed at the bottom of the experimental arena (Coulbourn) allowing for tracking the animal movements. If no movement was detected for 2 s the animals were considered to be freezing.

Context A and C were equipped with a wide-angle camera (IC capture) located underneath the platform allowing for the video tracking of orofacial movements (acquisition rate: 30 frames per second). Videos were later analyzed frame-by-frame and scoring of orofacial movements was performed manually. Hedonic orofacial movements were tongue protrusions, paw licking and licking or consumption of items in the arena. Aversive orofacial movements consist in gapes,

forelimbs flails, face wiping and chin rubbing. Neutral orofacial movements consist in low amplitude mouth movements and grooming. Analysis of the average bout duration for each type of hedonic orofacial movements revealed paw licking were often express by mice in continuous bouts. In order to normalize the contribution of each component to the hedonic taste reactivity score, paw lickings were then scored by time bins of 5s. This scoring method has been considered to be a more accurate measure of palatability.

Extracellular recordings in freely behaving mice

Prior to each behavioral session, electrodes were connected to a headstage (Plexon) containing 16 unity-gain operational amplifiers. The headstage was connected to a 16-channel computer-controlled preamplifier (gain 1000x, bandpass filter from 150 Hz to 9 kHz; Plexon). Neuronal activity was digitized at 40 kHz, bandpass filtered from 250 Hz to 8 kHz and isolated by time-amplitude window discrimination and template matching using a multichannel acquisition processor system (Plexon).

Single-unit spike sorting and analysis

Single-unit spike sorting was performed using Off-Line Spike Sorter (Plexon) as previously described[63]. Briefly, for each individual recording session, principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces and clusters containing similar waveforms were manually defined. A group of waveforms was considered to be generated from a single neuron if it defined a discrete cluster in the principal component space, distinct from clusters of other units and if it displayed refractory period of at least 1 ms. Multivariate analysis of variance (MANOVA), J3 statistic and Davis-Bouldin validity index (DB) were used to further confirm the sorting quality. Average waveforms of identified neurons were used to estimate the single-unit stability across recording sessions. Quantitative evaluation of the waveform shape similarity was assessed using linear correlation (r) values and only neurons displaying r values above 0.95 were considered as stable across sessions.

CONCLUSION

The work presented in this dissertation identifies the neuronal correlates of conditioning and extinction of opposite emotional valences. It demonstrates that excitatory and inhibitory learning are mostly encoded in a valence-specific manner in the basal amygdala. However, extinction of conditioned memories seems to rely on both valence-free and valence-specific mechanism.

This study is also the first investigation of the interaction between prior emotional experience and subsequent emotional associative learning. It shows that prior appetitive experience does not interfere with subsequent fear conditioning but leads to a deficit in fear extinction learning which correlates with a reduced activity in fear extinction neurons.

Finally, consistent with the dominance of aversive paradigms in the study of the cellular underpinnings of associative learning, we found amygdala activity to be strongly biased towards aversive events. This bias relies on the activity of extinction-resistant neurons, a discrete BA neuronal population participating in the maintenance over changes in CS-Us contingencies of valence-specific conditioned memories.

Taken together, our data demonstrates the high dimensionally of valence encoding and valence interaction both at the behavioral and neuronal levels.

REFERENCES

1. James, W. What is an emotion? *Mind* **9,** 188–205 (1884).
2. Dalgleish, T. The emotional brain. *Nat. Rev. Neurosci.* **5,** 583–589 (2004).
3. Lindquist, K. a. Emotions Emerge from More Basic Psychological Ingredients: A Modern Psychological Constructionist Model. *Emot. Rev.* **5,** 356–368 (2013).
4. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neuropsychiatry Clin. Neurosci.* **12,** 103–113 (1957).
5. Corkin, S. What's new with the amnesic patient H.M.? *Nat. Rev. Neurosci.* **3,** 153–60 (2002).
6. Liu, X. *et al.* Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **484,** 381–385 (2012).
7. Martin, S. J., Grimwood, P. D. & Morris, R. G. M. Synaptic plasticity and memory: An Evaluation of the Hypothesis. *Annu. Rev. Neurosci.* 649–711 (2000).
8. Whitlock, J. R., Heynen, A. J., Shuler, M. G. & Bear, M. F. Learning induces long-term potentiation in the hippocampus. *Science* **313,** 1093–1097 (2006).
9. Neves, G., Cooke, S. F. & Bliss, T. V. P. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nat. Rev. Neurosci.* **9,** 65–75 (2008).
10. Sah, P., Westbrook, R. F. & Lüthi, A. Fear conditioning and long-term potentiation in the amygdala: what really is the connection? *Ann. N. Y. Acad. Sci.* **1129,** 88–95 (2008).
11. Hübener, M. & Bonhoeffer, T. Searching for Engrams. *Neuron* **67,** 363–371 (2010).
12. Malenka, R. C. & Bear, M. F. LTP and LTD: an embarrassment of riches. *Neuron* **44,** 5–21 (2004).
13. Sjöström, P. J., Rancz, E. A., Roth, A. & Häusser, M. Dendritic excitability and synaptic plasticity. *Physiol. Rev.* **88,** 769–840 (2008).
14. Squire, L. R. & Zola-Morgan, S. The medial temporal lobe memory system. *Science* **253,** 1380–1386 (1991).
15. Bechara, A. *et al.* Double dissociation of conditioning and declarative knowledge relative to the amygdala and Hippocampus in Humans. **269,** 1115–1118 (1995).
16. Pavlov, I. P. Conditioned reflexes An Investigation of the Physiological Activity of the Cerebral Cortex. *Oxford Univ. Press* (1927).
17. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Class. Cond. II Curr. Res. theory* pp64-99 (1972).
18. Klüver, H. & Bucy, P. C. 'Psychic blindness' and other symptoms following bilateral temporal lobectomy in Rhesus monkeys. *Am. J. Physiol.* **Vol 119,** 352–353 (1937).
19. Appenzeller, S. *et al.* Amygdalae Calcifications Associated with Disease Duration in Lipoid Proteinosis. *J. Neuroimaging* **16,** 154–156 (2006).
20. Feinstein, J. S., Adolphs, R., Damasio, A. & Tranel, D. The Human Amygdala and the Induction and Experience of Fear. *Curr. Biol.* **21,** 34–38 (2011).
21. Lanteaume, L. *et al.* Emotion induction after direct intracerebral stimulations of human

amygdala. *Cereb. Cortex* **17,** 1307–1313 (2007).

22. Kemppainen, S. & Pitkänen, A. Distribution of parvalbumin, calretinin, and calbindin-D(28k) immunoreactivity in the rat amygdaloid complex and colocalization with gamma-aminobutyric acid. *J. Comp. Neurol.* **426,** 441–67 (2000).
23. Krettek, E. & Price, J. Amygdaloid projections to subcortical structures within the basal forebrain and brainstem in the rat and cat. *J Comp Neurol* 225–254 (1978).
24. McDonald, A. J. & Augustine, J. R. Localization of GABA-like immunoreactivity in the monkey amygdala. *Neuroscience* **52,** 281–294 (1993).
25. Carlsen, J. & Heimer, L. The basolateral amygdaloid complex as a cortical-like structure. *Brain Res.* **441,** 377–380 (1988).
26. Paré, D. & Smith, Y. Intrinsic circuitry of the amygdaloid complex: common principles of organization in rats and cats. *Trends Neurosci.* **21,** 240–241 (1998).
27. McDonald, A. . & Mascagni, F. Colocalization of calcium-binding proteins and GABA in neurons of the rat basolateral amygdala. *Neuroscience* **105,** 681–693 (2001).
28. Muller, J. F., Mascagni, F. & McDonald, A. J. Coupled networks of parvalbumin-immunoreactive interneurons in the rat basolateral amygdala. *J Neurosci* **25,** 7366–7376 (2005).
29. Millhouse, O. E. & DeOlmos, J. Neuronal configurations in lateral and basolateral amygdala. *Neuroscience* **10,** 1269–1300 (1983).
30. Lang, E. . & Paré, D. Synaptic responsiveness of interneurons of the cat lateral amygdaloid nucleus. *Neuroscience* **83,** 877–889 (1998).
31. Wolff, S. B. E. *et al.* Amygdala interneuron subtypes control fear learning through disinhibition. *Nature* **509,** 453–8 (2014).
32. Jolkkonen, E. & Pitkänen, A. Intrinsic connections of the rat amygdaloid complex: projections originating in the central nucleus. *J. Comp. Neurol.* **395,** 53–72 (1998).
33. Sun, N. & Cassell, M. D. Intrinsic GABAergic neurons in the rat central extended amygdala. *The Journal of comparative neurology* 381–404 (1993).
34. Luskin, M. B. & Price, J. L. The Laminar Distribution of Intracortical Fibers Originating in the Olfactory Cortex of the Rat. *The Journal of comparative neurology* 292–302 (2004).
35. LeDoux, J. E., Farb, C. & Ruggiero, D. a. Topographic organization of neurons in the acoustic thalamus that project to the amygdala. *J. Neurosci.* **10,** 1043–1054 (1990).
36. LeDoux, J. E., Farb, C. R. & Romanski, L. M. Overlapping projections to the amygdala and striatum from auditory processing areas of the thalamus and cortex. *Neurosci. Lett.* **134,** 139–144 (1991).
37. Shi, C. J. & Cassell, M. D. Cascade projections from somatosensory cortex to the rat basolateral amygdala via the parietal insular cortex. *J. Comp. Neurol.* **399,** 469–91 (1998).
38. Shi, C. & Davis, M. Visual pathways involved in fear conditioning measured with fear-potentiated startle: behavioral and anatomic studies. *J. Neurosci.* **21,** 9844–55 (2001).
39. Amaral, D. G. & Insausti, R. Retrograde transport of D-[3H]-aspartate injected into the

monkey amygdaloid complex. *Exp. Brain Res.* **88,** 375–388 (1992).

40. Campeau, S. & Davis, M. Involvement of subcortical and cortical afferents to the lateral nucleus of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *J. Neurosci.* **15,** 2312–27 (1995).

41. Mcdonald, A. J., Mascagni, F. & Guo, L. Projections of the medial and lateral prefrontal cortices to the amygdala: a Phaseolus vulgaris leucoagglutinin study in the rat. *Neuroscience* **71,** 55–75 (1996).

42. McDonald, A. J. Cortical pathways to the mammalian amygdala. *Prog. Neurobiol.* **55,** 257–332 (1998).

43. Pitkänen, A., Pikkarainen, M., Nurminen, N. & Ylinen, A. Reciprocal Connections between the Amygdala and the Hippocampal Formation, Perirhinal Cortex, and Postrhinal Cortex in Rat: A Review. *Ann. N. Y. Acad. Sci.* **911,** 369–391 (2006).

44. Corcoran, K. A. & Maren, S. Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. *J. Neurosci.* **21,** 1720–6 (2001).

45. Corcoran, K. A., Desmond, T. J., Frey, K. A. & Maren, S. Hippocampal inactivation disrupts the acquisition and contextual encoding of fear extinction. *J. Neurosci.* **25,** 8978–87 (2005).

46. Sotres-Bayon, F., Bush, D. E. A. & LeDoux, J. E. Emotional Perseveration: An Update on Prefrontal-Amygdala Interactions in Fear Extinction. *Learn. Mem.* **11,** 525–535 (2004).

47. Stalnaker, T. A. *et al.* Cocaine-induced decision-making deficits are mediated by miscoding in basolateral amygdala. *Nat Neurosci* **10,** 949–951 (2007).

48. Duvarci, S. & Pare, D. Amygdala microcircuits controlling learned fear. *Neuron* **82,** 966–980 (2014).

49. Killcross, S., Robbins, T. W. & Everitt, B. J. Different types of fearconditioned behaviour mediated by separate nuclei within amygdala. *Nature* 377–380 (1997).

50. Parkinson, J. a., Robbins, T. W. & Everitt, B. J. Dissociable roles of the central and basolateral amygdala in appetitive emotional learning. *Eur. J. Neurosci.* **12,** 405–413 (2000).

51. Weiskrantz, L. Behavioral changes associated with ablation of the amygdaloid complex in monkeys. *The Journal of comparative and physiological psychology* (1956).

52. Delgado, J. M., Rosvold, H. E. & Looney, E. Evoking conditioned fear by electrical stimulation of subcortical structures in the monkey brain. *The Journal of comparative and physiological psychology* 373–380 (1956).

53. Cousens, G. & Otto, T. Both pre- and posttraining excitotoxic lesions of the basolateral amygdala abolish the expression of olfactory and contextual fear conditioning. *Behav. Neurosci.* **112,** 1092–103 (1998).

54. Wilensky, a E., Schafe, G. E. & LeDoux, J. E. Functional inactivation of the amygdala before but not after auditory fear conditioning prevents memory formation. *J. Neurosci.* **19,** RC48 (1999).

55. Johansen, J. P. *et al.* Optical activation of lateral amygdala pyramidal cells instructs associative fear learning. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12692–12697 (2010).

56. Buchel, C., Morris, J., Dolan, R. J. & Friston, K. J. Brain Systems Mediating Aversive Conditioning : an Event-Related fMRI Study. **20,** 947–957 (1998).

57. Rogan, M. T. & LeDoux, J. E. LTP is accompanied by commensurate enhancement of auditory-evoked responses in a fear conditioning circuit. *Neuron* **15,** 127–136 (1995).

58. Collins, D. R. & Paré, D. Differential Fear Conditioning Induces Reciprocal Changes in the Sensory Responses of Lateral Amygdala Neurons to the CS+ and CS–. *Learn. Mem.* **7,** 97–103 (2000).

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

60. Quirk, G. J., Armony, J. L. & Ledoux, J. E. Components of Tone-Evoked Spike Trains in Auditory Cortex and Lateral Amygdala. *Neuron* **19,** 613–624 (1997).

61. Repa, J. C. *et al.* Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nat. Neurosci.* **4,** 724–731 (2001).

62. An, B., Hong, I. & Choi, S. Long-Term Neural Correlates of Reversible Fear Learning in the Lateral Amygdala. *J. Neurosci.* **32,** 16845–16856 (2012).

63. Herry, C. *et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454,** 600–606 (2008).

64. Rogan, M. T., Stäubli, U. V & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390,** 604–7 (1997).

65. Walker, D. L. & Davis, M. Involvement of NMDA receptors within the amygdala in short-versus long-term memory for fear conditioning as assessed with fear-potentiated startle. *Behav. Neurosci.* **114,** 1019–1033 (2000).

66. Maren, S. Synaptic mechanisms of associative memory in the amygdala. *Neuron* **47,** 783–786 (2005).

67. Hatfield, T., Han, J. S., Conley, M., Gallagher, M. & Holland, P. Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J. Neurosci.* **16,** 5256–65 (1996).

68. Gottfried, J. A., O'Doherty, J. & Dolan, R. J. Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J. Neurosci.* **22,** 10829–37 (2002).

69. O'Doherty, J. P., Deichmann, R., Critchley, H. D. & Dolan, R. J. Neural Responses during Anticipation of a Primary Taste Reward. *Neuron* **33,** 815–826 (2002).

70. Bermudez, M. A. & Schultz, W. Reward Magnitude Coding in Primate Amygdala Neurons. *J. Neurophysiol.* **104,** 3424–3432 (2010).

71. Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439,** 865–870 (2006).

72. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* **26,** 321–52 (2002).

73. Martin-Soelch, C., Linthicum, J. & Ernst, M. Appetitive conditioning: Neural bases and implications for psychopathology. *Neurosci. Biobehav. Rev.* **31,** 426–440 (2007).

74. Ambroggi, F., Ishikawa, A., Fields, H. L. & Nicola, S. M. Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron* **59,** 648–661 (2008).

75. Stuber, G. D. *et al.* Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature* **475,** 377–380 (2011).

76. Bouton, M. E., Westbrook, R. F., Corcoran, K. A. & Maren, S. Contextual and temporal modulation of extinction: behavioral and biological mechanisms. *Biol Psychiatry* **60,** 352–360 (2006).

77. Myers, K. M. & Davis, M. Mechanisms of fear extinction. *Mol. Psychiatry* **12,** 120–150 (2007).

78. Herry, C. *et al.* Neuronal circuits of fear extinction. *Eur J Neurosci* **31,** 599–612 (2010).

79. Falls, W. A., Miserendino, M. J. & Davis, M. Extinction of fear-potentiated startle: blockade by infusion of an NMDA antagonist into the amygdala. *J. Neurosci.* **12,** 854–63 (1992).

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

81. Senn, V. *et al.* Long-range connectivity defines behavioral specificity of amygdala neurons. *Neuron* **81,** 428–437 (2014).

82. Bouton, M. E. Learning and the persistence of appetite: Extinction and the motivation to eat and overeat. *Physiol. Behav.* **103,** 51–58 (2011).

83. Burns, L. H., Everitt, B. J. & Robbins, T. W. Effects of excitotoxic lesions of the basolateral amygdala on conditional discrimination learning with primary and conditioned reinforcement. *Behav. Brain Res.* **100,** 123–33 (1999).

84. McLaughlin, R. J. & Floresco, S. B. The role of different subregions of the basolateral amygdala in cue-induced reinstatement and extinction of food-seeking behavior. *Neuroscience* **146,** 1484–1494 (2007).

85. Fuchs, R. A., Feltenstein, M. W. & See, R. E. The role of the basolateral amygdala in stimulus-reward memory and extinction memory consolidation and in subsequent conditioned cued reinstatement of cocaine seeking. *Eur J Neurosci* **23,** 2809–2813 (2006).

86. Rhodes, S. E. V & Killcross, a. S. Lesions of rat infralimbic cortex enhance renewal of extinguished appetitive Pavlovian responding. *Eur. J. Neurosci.* **25,** 2498–2503 (2007).

87. Ernst, M. *et al.* Amygdala and nucleus accumbens in responses to receipt and omission of gains in adults and adolescents. *Neuroimage* **25,** 1279–1291 (2005).

88. Belova, M. A., Paton, J. J., Morrison, S. E. & Salzman, C. D. Expectation modulates neural

responses to pleasant and aversive stimuli in primate amygdala. *Neuron* **55,** 970–984 (2007).

89. Tye, K. M. & Janak, P. H. Amygdala neurons differentially encode motivation and reinforcement. *J Neurosci* **27,** 3937–3945 (2007).

90. Bouton, M. E. & Peck, C. E. Spontaneous recovery in cross-motivational transfer. *Anim. Learn. Behav.* **20,** 313–321 (1992).

91. Nasser, H. M. & McNally, G. P. Appetitive-aversive interactions in Pavlovian fear conditioning. *Behav Neurosci* **126,** 404–422 (2012).

92. Nasser, H. M. & McNally, G. P. Neural correlates of appetitive-aversive interactions in Pavlovian fear conditioning. *Learn Mem* **20,** 220–228 (2013).

93. Srebro, B. Looking back at Jerzy Konorski's book 'integrative activity of the brain', 45 years after. *Acta Neurobiol. Exp. (Wars).* **73,** 451–462 (2013).

94. Konorski, J. *Integrative activity of the brain: an interdisciplinary approach*. (1967).

95. Xiu, J. *et al.* Visualizing an emotional valence map in the limbic forebrain by TAI-FISH. *Nat. Neurosci.* **17,** 1552–1559 (2014).

96. Herry, C. *et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454,** 600–606 (2008).

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

98. Collins, D. R. & Paré, D. Differential Fear Conditioning Induces Reciprocal Changes in the Sensory Responses of Lateral Amygdala Neurons to the CS+ and CS-. *Learn. Mem.* **7,** 97–103 (2000).

99. Rosenkranz, J. A. & Grace, A. A. Dopamine-mediated modulation of odour-evoked amygdala potentials during pavlovian conditioning. *Nature* **417,** 282–287 (2002).

100. Goosens, K. A. *et al.* Auditory-evoked spike firing in the lateral amygdala and Pavlovian fear conditioning: mnemonic code or fear bias? *Neuron* **40,** 1013–22 (2003).

101. LeDoux, J. E. Emotion Circuits in the Brain. *Annu. Rev. Neurosci.* **23,** 155–184 (2000).

102. Maren, S. & Quirk, G. J. Neuronal signalling of fear memory. *Nat. Rev. Neurosci.* **5,** 844–852 (2004).

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

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

105. Rosenkranz, J. A., Moore, H. & Grace, A. A. The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. *J. Neurosci.* **23,** 11054–64 (2003).

106. Berretta, S., Pantazopoulos, H., Caldera, M., Pantazopoulos, P. & Paré, D. Infralimbic cortex activation increases c-fos expression in intercalated neurons of the amygdala.

Neuroscience **132,** 943–953 (2005).

107. Pare, D., Quirk, G. J. & Ledoux, J. E. New Vistas on Amygdala Networks in Conditioned Fear. *J. Neurophysiol.* **92,** 1–9 (2004).

108. Lu, K. T., Walker, D. L. & Davis, M. Mitogen-activated protein kinase cascade in the basolateral nucleus of amygdala is involved in extinction of fear-potentiated startle. *J. Neurosci.* **21,** RC162 (2001).

109. Herry, C., Trifilieff, P., Micheau, J., Lüthi, A. & Mons, N. Extinction of auditory fear conditioning requires MAPK/ERK activation in the basolateral amygdala. *Eur. J. Neurosci.* **24,** 261–269 (2006).

110. Sotres-Bayon, F., Bush, D. E. A. & LeDoux, J. E. Acquisition of Fear Extinction Requires Activation of NR2B-Containing NMDA Receptors in the Lateral Amygdala. *Neuropsychopharmacology* **32,** 1929–1940 (2007).

111. Quirk, G. J., Garcia, R. & González-Lima, F. Prefrontal Mechanisms in Extinction of Conditioned Fear. *Biol. Psychiatry* **60,** 337–343 (2006).

112. Bouton, M. E. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol. Psychiatry* **52,** 976–86 (2002).

113. Pitkänen, A., Pikkarainen, M., Nurminen, N. & Ylinen, A. Reciprocal connections between the amygdala and the hippocampal formation, perirhinal cortex, and postrhinal cortex in rat. A review. *Ann. N. Y. Acad. Sci.* **911,** 369–91 (2000).

114. Hobin, J. A., Ji, J. & Maren, S. Ventral hippocampal muscimol disrupts context-specific fear memory retrieval after extinction in rats. *Hippocampus* **16,** 174–182 (2006).

115. McDonald, A. J. Organization of amygdaloid projections to the prefrontal cortex and associated striatum in the rat. *Neuroscience* **44,** 1–14 (1991).

116. Herry, C. & Mons, N. Resistance to extinction is associated with impaired immediate early gene induction in medial prefrontal cortex and amygdala. *Eur. J. Neurosci.* **20,** 781–790 (2004).

117. Muramoto, K., Ono, T., Nishijo, H. & Fukuda, M. Rat amygdaloid neuron responses during auditory discrimination. *Neuroscience* **52,** 621–36 (1993).

118. Maren, S., Poremba, A. & Gabriel, M. Basolateral amygdaloid multi-unit neuronal correlates of discriminative avoidance learning in rabbits. *Brain Res.* **549,** 311–6 (1991).

119. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: Implications of a quantitative analysis. *Proc. Natl. Acad. Sci.* **101,** 13124–13131 (2004).

120. Herry, C. & Garcia, R. Prefrontal Cortex Long-Term Potentiation, But Not Long-Term Depression, Is Associated with the Maintenance of Extinction of Learned Fear in Mice. *J. Neurosci.* **22,** (2002).

121. Likhtik, E., Pelletier, J. G., Popescu, A. T. & Pare, D. Identification of Basolateral Amygdala Projection Cells and Interneurons Using Extracellular Recordings. *J. Neurophysiol.* **96,** 3257–3265 (2006).

122. Stalnaker, T. A., Franz, T. M., Singh, T. & Schoenbaum, G. Basolateral Amygdala Lesions Abolish Orbitofrontal-Dependent Reversal Impairments. *Neuron* **54,** 51–58 (2007).

123. Stalnaker, T. A. *et al.* Cocaine-induced decision-making deficits are mediated by miscoding in basolateral amygdala. *Nat. Neurosci.* **10,** 949–951 (2007).

124. Paré, D. Role of the basolateral amygdala in memory consolidation. *Prog. Neurobiol.* **70,** 409–20 (2003).

125. McGaugh, J. L. The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci.* **27,** 1–28 (2004).

126. Paz, R., Pelletier, J. G., Bauer, E. P. & Paré, D. Emotional enhancement of memory via amygdala-driven facilitation of rhinal interactions. *Nat. Neurosci.* **9,** 1321–1329 (2006).

127. Nader, K., Amorapanth, P. & LeDoux, J. E. Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nat. Neurosci.* **3,** 74–79 (2000).

128. Nader, K., Majidishad, P., Amorapanth, P. & LeDoux, J. E. Damage to the lateral and central, but not other, amygdaloid nuclei prevents the acquisition of auditory fear conditioning. *Learn. Mem.* **8,** 156–63 (2001).

129. Anglada-Figueroa, D. & Quirk, G. J. Lesions of the Basal Amygdala Block Expression of Conditioned Fear But Not Extinction. *J. Neurosci.* **25,** 9680–9685 (2005).

130. Goosens, K. A. & Maren, S. Contextual and Auditory Fear Conditioning are Mediated by the Lateral, Basal, and Central Amygdaloid Nuclei in Rats. *Learn. Mem.* **8,** 148–155 (2001).

131. Jing, J. & Gillette, R. Escape swim network interneurons have diverse roles in behavioral switching and putative arousal in Pleurobranchaea. *J. Neurophysiol.* **83,** 1346–55 (2000).

132. Yapici, N., Kim, Y.-J., Ribeiro, C. & Dickson, B. J. A receptor that mediates the post-mating switch in Drosophila reproductive behaviour. *Nature* **451,** 33–37 (2008).

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

134. Rodriguez, B. I., Craske, M. G., Mineka, S. & Hladek, D. Context-specificity of relapse: effects of therapist and environmental context on return of fear. *Behav. Res. Ther.* **37,** 845–62 (1999).

135. Paxinos, G. & Franklin, K. B. J. *The mouse brain in stereotaxic coordinates*. (Elsevier Academic Press, 2004).

136. Herry, C. *et al.* Processing of Temporal Unpredictability in Human and Animal Amygdala. *J. Neurosci.* **27,** 5958–5966 (2007).

137. Nicolelis, M. A. L. *et al.* Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc. Natl. Acad. Sci.* **100,** 11041–11046 (2003).

138. Jackson, A. & Fetz, E. E. Compact Movable Microwire Array for Long-Term Chronic Unit Recording in Cerebral Cortex of Primates. *J. Neurophysiol.* **98,** 3109–3118 (2007).

139. Grill, H. J. & Norgren, R. The taste reactivity test. I. Mimetic responses to gustatory stimuli in neurologically normal rats. *Brain Res.* **143,** 263–79 (1978).

140. Cagniard, B. & Murphy, N. P. Taste reactivity and its modulation by morphine and methamphetamine in C57BL/6 and DBA/2 mice. *Physiol Behav* **96,** 412–420 (2009).

141. Wheeler, R. A. *et al.* Behavioral and electrophysiological indices of negative affect predict cocaine self-administration. *Neuron* **57,** 774–785 (2008).

142. Lin, J.-Y., Arthurs, J. & Reilly, S. Conditioned taste aversion, drugs of abuse and palatability. *Neurosci. Biobehav. Rev.* **45,** 28–45 (2014).

143. West, E. A. *et al.* Transient inactivation of basolateral amygdala during selective satiation disrupts reinforcer devaluation in rats. *Behav. Neurosci.* **126,** 563–74 (2012).

144. Robinson, M. J. F. & Berridge, K. C. Instant transformation of learned repulsion into motivational 'wanting'. *Curr. Biol.* **23,** 282–9 (2013).

145. Uwano, T., Nishijo, H., Ono, T. & Tamura, R. Neuronal responsiveness to various sensory stimuli, and associative learning in the rat amygdala. *Neuroscience* **68,** 339–361 (1995).

146. Shabel, S. J., Schairer, W., Donahue, R. J., Powell, V. & Janak, P. H. Similar neural activity during fear and disgust in the rat basolateral amygdala. *PLoS One* **6,** e27797 (2011).

147. Namburi, P. *et al.* A circuit mechanism for differentiating positive and negative associations. *Nature* **520,** 675–678 (2015).

148. Málková, L., Gaffan, D. & Murray, E. A. Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. *J. Neurosci.* **17,** 6011–20 (1997).

149. De Martino, B., Kumaran, D., Seymour, B. & Dolan, R. J. Frames, biases, and rational decision-making in the human brain. *Science (80-.).* **313,** 684–687 (2006).

150. De Martino, B., Camerer, C. F. & Adolphs, R. Amygdala damage eliminates monetary loss aversion. *Proc Natl Acad Sci U S A* **107,** 3788–3792 (2010).

151. Carelli, R. M. & West, E. A. When a good taste turns bad: Neural mechanisms underlying the emergence of negative affect and associated natural reward devaluation by cocaine. *Neuropharmacology* **76 Pt B,** 360–9 (2014).

152. Hardin, M. G., Pine, D. S. & Ernst, M. The influence of context valence in the neural coding of monetary outcomes. *Neuroimage* **48,** 249–257 (2009).

153. Mesquita, B., Feldman Barrett, L. & Smith, E. R. The mind in context. *Guilford Press* (2010).

ACKNOWLEDGMENTS

I would like to thank my Ph.D. supervisor Andreas Lüthi for giving me the opportunity to work on this fascinating project, for his support and mentoring along my Ph.D. and for such a fantastic time in his lab.

I am extremely thankful to Cyril Herry for introducing me to single-unit recordings and in vivo experiments. I wanted to study emotions at the single spike level since the very beginning of my studies, so many thanks to him for making it become reality.

I also remember with a lot of gratitude Stéphane Ciocchi's patience during my first spike sorting sessions.

I would also like to thank the members of my Ph.D. committee, Rainer Friedrich and Kaspar Vogt, for their extremely valuable advice during the thesis committee meetings and for their interest and support of my Ph.D. work.

Thanks to Michael Stadler for his priceless help with R-scripts which allowed me to handle the size and the multidimensionality of my dataset.

Thanks to Sjoerd Van Eeden, Alan Naylor and Thomas Nyffenegger from the IT department for saving me countless times from nightmarish bugs and crashes.

I would like to thank as well Steven Bourke and Laurent Gelman from the imaging facility for their help with microscopy.

Many thanks to Paul Argast and Peter Buchmann from the micromechanics and electronic workshops. Their help was absolutely instrumental in building high-quality recordings contexts and surgical implants.

I would like to thank Christian Müller for his help with surgeries and implant building.

Thanks to my colleagues, Julien Courtin, Nigel Whittle, Tingjia Lu and Sol Fustiñana for proofreading my Ph.D. dissertation.

An enormous thank to Milica Markovic, she knows why ;)

Finally, thanks to my current and former colleagues from the Lüthi lab, for their help and support, for their availability and great advice, and for all the good times we shared together.