STATE-DEPENDENT REGULATION OF FEAR EXTINCTION LEARNING BY THE INTEROCEPTIVE INSULAR CORTEX

Alexandra S. Klein

Dissertation

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

München, den 29.06.2020

Betreuerin: Dr. Nadine Gogolla
Erstgutachter: Prof. Dr. Rüdiger Klein
Zweitgutachter: Prof. Dr. Laura Busse

Tag der Einreichung: 29.06.2020
Tag der mündlichen Prüfung: 18.12.2020

Die Arbeit, die in dieser Dissertation vorliegt, wurde durchgeführt im Labor von Dr. Nadine Gogolla, Forschungsgruppe ‘Circuits for Emotions’, Max-Planck-Institut für Neurobiologie, Martinsried.

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 29.06.2020 Alexandra Klein

(Unterschrift)

Erklärung

Hiermit erkläre ich, *

☒ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

☒ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.

☐ ~~dass ich mich mit Erfolg der Doktorprüfung im Hauptfach ..... und in den Nebenfächern ..... bei der Fakultät für ..... der ..... (Hochschule/Universität) unterzogen habe.~~

☐ ~~dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.~~

München, den 29.06.2020 Alexandra Klein

(Unterschrift)

*) Nichtzutreffendes streichen

ABSTRACT

Fear is an adaptive behavioral response to avoid or reduce harm and thus ensure survival. However, fear is maladaptive when it persists in the absence of direct threat or when it prevents an organism to attend to its homeostatic needs. The interoceptive, posterior insular cortex (pIC) is the major cortical recipient for signals from inside the body and is at the same time increasingly recognized as an intricate part of a wider neuronal network regulating fear and anxiety. However, the neuronal underpinnings of an insular role in fear regulation remain elusive.

In my thesis, I demonstrated that the pIC regulates the extinction of learned fear in a state dependent manner.

Using fiber photometry recordings of excitatory neuron activity and electrophysiological single-unit recordings in the mouse pIC, I found that the pIC processes state-dependently fear-eliciting cues and painful stimuli. Further, I revealed that specific subpopulations of pIC neurons encoded sustained affective states.

I next conducted optogenetic silencing experiments to decipher the involvement of the pIC in extinction learning and revealed surprising, bidirectional effects: while pIC inhibition facilitated extinction learning in animals displaying a low internal fear state, the same circuit manipulation impaired fear extinction learning in animals displaying a high internal fear state. Additionally, in an attempt to unravel how internal states can influence pIC activity, I developed a vagus nerve stimulation (VNS) protocol to manipulate interoceptive processes during extinction. I found that stimulation of the vagus nerve, that carries bodily signals to the brain, elicited insula activity and regulated extinction learning in the same, bidirectional and state-dependent manner.

Taken together, my data suggest that the pIC maintains optimal fear extinction learning despite differences in internal fear states, a process that may be modulated by vagal afferents transmitting interoceptive signals to the brain.

ACKNOWLEDGEMENTS

First of all, I want to say a gigantic “Thank You” to Nadine Gogolla for being an amazing supervisor and mentor, all along the way (more than 6 years now). Thanks for your trust to choose me as one of your first PhD students, it was such a great and educational experience to help setting up a new lab from scratch. Thanks for always being supportive and giving me so many opportunities to learn and grow as a scientist. I enjoyed each step of the journey, all ups and downs included, and I am really proud of what the group has already achieved - but there is for sure more to come.

Next, I want to thank my thesis advisory committee members Ruben Portugues and Johannes Letzkus for their extremely valuable inputs and their support during the last six years. Each single TAC meeting helped me to steer my project into a better direction.

I am extremely thankful that I was working with such amazing, smart, and supportive colleagues from the Gogolla lab. You are the reason why “going to work” more feels like “going to meet friends”. Thanks for your inputs and our hour-long discussion over coffee (or beer). Thanks for being great people - inside and outside of the lab! In detail, I want to thank Daniel for being at my side from day 1 of the PhD, Nate for never giving up on trying to find an explanation for my always confusing data, Caro for being the BEST master student a final year PhD can hope for, Alja for always being such a chill dude who is able to calm me down in stressful situations, Tom B. for always having an open ear for me, Meryl for completing the lab and being not only a great colleague but also a friend, Jeong for being so helpful with ePhys data analysis, Eunjae for cutting countless brains, Frederique for all the mouse husbandry, Tom G. for being a great desk partner who made every day lab life much more fun, Onur for finally getting DLC to work and my former Master student Arthur who did all the seminal work on global pIC manipulations. Thanks to the whole Klein lab for always helping us out with materials / knowledge / technical assistance / coffee.

Most importantly, I need to thank my friends and family, who were standing by my side throughout the last (sometimes very difficult) last 6 years. Thanks to all my parents Gabi, Stefan and Peter for supporting each of my decisions throughout my life. I am so grateful that you always believed in me to be the first one in our family to follow an academic career path. Thanks to my best friends Hugo, Kaspar, Resi, Nina and my Flamongos Hita and Toni, for always being there for me, listening to my worries and struggles and making me feel better in an instance. And most of all, thanks to Flo, who was the ABSOLUTELY BEST support I could have ever wished for, especially now during lockdown and thesis writing times - thanks for all the tasty lunches, dinners, coffees and treats that kept my mind sane! Even though it was not always easy for you (long-distance relationship, long hours at work, commuting, etc.), you always stood by my side and supported me and my scientific career. I will forever be grateful for that.

CONTENTS

ABBREVIATIONS AND ACRONYMS

AAV	Adeno-associated virus
AI	Agranular insular cortex
aIC	Anterior insular cortex
AID	Agranular insular cortex, dorsal part
AIP	Agranular insular cortex, posterior part
AIV	Agranular insular cortex, ventral part
AP	Anterior-posterior
BW	Body weight
CaMKII	Ca2+/calmodulin-dependent protein kinase
CeA	Central amygdala
ChR2	Channelrhodpsin
CS-	Conditioned stimulus (not paired with US)
CS+	Conditioned stimulus (paired with US)
DI	Dysgranular insular cortex
E, E early, E mid, E late	Extinction (early, mid, late)
eNpHR3.0	enhanced Natronomonas pharaonis Halorhodopsin 3.0
EPM	Elevated-plus maze
ER	Extinction Retrieval
eYFP	Enhanced yellow fluorescent protein
FC, weak FC, strong FC	Fear conditioning (weak, strong)
FC/Ext	Fear conditioning and extinction paradigm
fMRI	Functional magnetic resonance imaging
FP	Fiber photometry

GECIs	Genetically encoded calcium indicators
GI	Granular insular cortex
H	Habituation
IC	Insular cortex
LC	Locus coeruleus
LH	Lateral hypothalamus
NAc	Nucleus accumbens core
NTS	Nucleus of solitary tract
OF	Open field test
PBN	Parabrachial nucleus
pIC	posterior insular cortex
preCS	30 seconds preceding a CS
PTSD	Post-traumatic stress disorder
R	Recall
SDA	subdiaphragamactic vagal deafferentiation
US	Unconditioned stimulus
VNS	Vagus nerve stimulation
VPM	Ventral posteriomedial nucleus of the thalamus

1 INTRODUCTION

1.1 Fear and anxiety

Fear induces behavioral and physiological responses aimed at avoiding or reducing harm and thus ensuring survival. However, fear is maladaptive when it persists in the absence of threat, which represents a hallmark of anxiety disorders. Importantly, 'anxiety disorder' is a comprehensive term for pathological and inappropriate expression of fear. According to the DSM-IV and ICD-10 classifications, anxiety disorders include, among others, panic disorders, social or specific phobia, agoraphobia, obsessive–compulsive disorder, generalized anxiety disorder, as well as somatoform disorders[1]. Anxiety disorders are the most common psychiatric disorders occurring in up to one-third of the population in their lifetime [2], thus impacting not only affected individuals, but also their relatives and the society. Therefore, there is an imperative need for effective therapies for anxiety disorders.

1.1.1 Therapies for anxiety disorders

As the various forms of anxiety disorders differ in their etiology, there is no universally applicable treatment. However, the prevailing standard approach for treating most types of anxiety disorders consists of either medication and / or psychological therapy. Several meta-analyses have proven the efficacy of pharmacotherapy [1,3–5], however, as drugs commonly evoke severe side effects, there is a growing focus on implementing psychological therapies. Indeed, clinical guidelines suggest that trauma-focused psychotherapy is more effective and should be the preferred initial treatment [6,7]. One of the most commonly used approaches of trauma-focused psychotherapy is 'exposure therapy', defined as any treatment that encourages the systematic confrontation with the feared stimulus. Those stimuli can be either external (e.g. fear-evoking objects, situations or activities) or internal (e.g. fear-evoking thoughts

or physical sensations) [8,9]. The overall aim of this therapy is to inhibit the patient's fearful reaction to the stimulus by re-exposing the individual to the fear eliciting stimuli. However, exposure therapy is usually conducted in the safe environment of a clinic or doctor's office, or under pharmacological treatment that influences the patient's internal state. It has been shown that even when fear responses are successfully reduced in these settings, it may be difficult to transfer this novel memory to other contexts[10]. Thus, fearful reactions often reemerge when the patients are reintegrated in their everyday life and/or when medication is removed. This is due to a phenomenon called 'state-dependence' of memory, describing that recall of memory is enhanced if external context and/or internal-state is the same at time of encoding and time of recall [11].

Consequently, a better understanding of the brain mechanisms involved in the successful inhibition of fear and anxiety has become an enormous research field in the scientific community in recent year.

1.1.2 Fear extinction as an approach to study fear inhibition in the laboratory

To study fear inhibition in human and rodents in the laboratory, a procedure call 'fear extinction' that is based on the principles of exposure therapy has been in practice since the early 20th century [12].

In general, the fundamental idea of fear extinction is the following: a subject who previously learned that a stimulus predicts an aversive event is repeatedly exposed to that stimulus in the absence of any aversive event. If fear extinction is successful, the stimulus does not evoke any fear responses in the subject anymore[13]. To implement this in laboratory settings, the test subject (e.g. human, rat or mouse) first undergoes a classic Pavlovian fear conditioning paradigm (**Figure 1**), where an initially neutral stimulus (the conditioned stimulus or CS; for example, a light or a sound) is repeatedly paired with an aversive outcome (the unconditioned stimulus or US; for example a mild electric shock or a loud noise burst). Thus, the subject learns the predictive ability of the CS as to the occurrence of the US, ultimately exhibiting a conditioned fear response to the CS (**Figure 1**, blue curve). In humans, this response is usually measured by an increase in skin-conductance, whereas in rodents, the classic read-out of fear behavior is 'freezing', defined as the cessation of any movement (except breathing). The fear conditioning success is usually tested in a recall session, where the fear-evoking CS is played only a few times without any aversive outcome and the conditioned fear responses are measured. During extinction training, the fear-evoking CS is repeatedly presented in the absence of the US, so that over time, the predictive ability of the CS is gradually reduced. Consequently, also the conditioned fear response declines over the course of extinction training (**Figure 1**, good within session extinction (blue curve)), resulting in diminished or in

the optimal case even completely abolished fear responses to the fear-evoking stimulus (**Figure 1.**, good extinction retrieval (green data point)).

Importantly, extinction is not simply causing 'forgetting' (i.e. erasure or disruption) of the conditioned fear memory. First, numerous studies have shown that extinction requires exposure to the CS in the absence of the US as opposed to the simple passage of time [14]. Therefore, extinction is rather an active learning process that requires additional training [15]. Second, the process of extinction learning is not always successful, and extinction memory is inherently fragile (**Figure 1**). For example, in patients suffering from anxiety- or trauma-related disorders, these inhibitory learning mechanisms are often dysfunctional and thus, poor extinction learning is performed that only minimally reduces conditioned fear responses (**Figure 1**, poor within session extinction (red curve) and poor extinction retrieval, orange data point). Further, even after initially successful extinction learning (**Figure 1**, good within session extinction (blue curve)), relapse of the fear reaction can occur (poor extinction retrieval, orange data point). Relapses of fear reactions can happen through processes like renewal, reinstatement or even spontaneously [16,17].

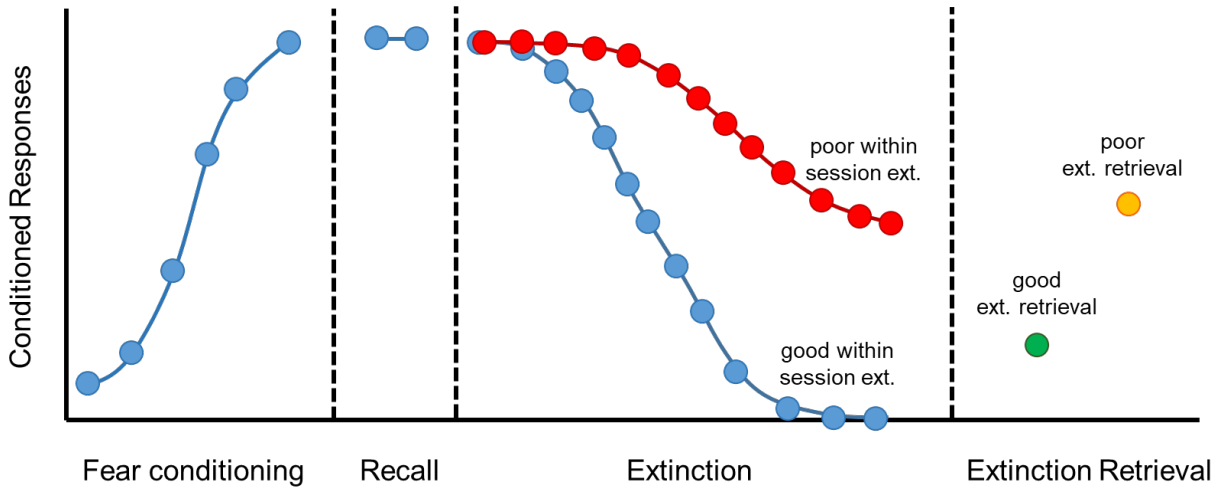


Figure 1: Process of fear and extinction learning.
During fear conditioning, conditioned responses to the CS increase by being paired to an aversive US and the success of fear conditioning is evaluated in a recall session, where the CS is presented only a few times, but without being paired to the US. Good within session extinction is characterized by a decrease in conditioned responses upon several presentation of the CS without the US, whereas poor within session extinction causes only a minimal reduction of the conditioned response. Extinction retrieval happens at a later time point, when the CS is again presented. Good extinction retrieval is characterized by low levels of conditioned responses (green data point), whereas poor extinction retrieval is characterized by high levels of conditioned responses (orange data point). Figure adapted from [17].

Hence, this indicates that the process of extinction generates a new 'inhibitory memory' that competes with the expression of the fearful conditioned memory [18]. The neuronal mechanisms of fear extinction thus seem extremely complex and must include the coordinated communication of numerous brain regions that are involved in the correct formation, consolidation, storage and retrieval of extinction memories. Even though neuroscientific research has made substantial progress in identifying brain regions involved in fear condition and extinction in the last decades [13,16,19,20], there is still an urgent need to better understand the neuronal mechanisms underlying fear extinction as this might help to inform and improve the efficacy of these forms of treatment.

1.2 The Insular cortex

Human imaging studies have linked the IC with multiple complex functions, such as sensorimotor, olfactory-gustatory, cognitive as well as socio-emotional processes [21]. More recent studies have identified the IC as a core region affected across many psychiatric and neurological disorders[22], highlighting the role of the IC in emotion processing and regulation [23,24]. In particular, the IC is suggested to play a pivotal role in different forms of anxiety disorders [25]. Moreover, there has been a surge of interest in the human IC as it is the core region involved in the processing of bodily signals, also called interoception [26,27]. However, despite this overwhelming evidence for more complex functions, animal studies have mostly addressed the IC's involvement in taste processing [28–34].

Our laboratory hypothesized that the insula serves as an interface between bodily and emotional states [35], thus placing it in a pivotal position for regulating fear behavior. However, how this is mechanistically implemented is currently not known. Thus, in my thesis I wanted to study the role of the IC in fear and anxiety-related behaviors.

In the following, I will briefly review the current state of knowledge about the anatomical organization and function of the rodent IC and introduce the main techniques that I used in this thesis.

1.2.1 Anatomical organization of the mouse insular cortex

In primates, including humans, the IC is a deeply invaginated cortical area and is located in the depth of the Sylvian fissure. It is therefore completely hidden by the temporal, frontal and parietal lobes of the brain (**Figure 2**, left). In the lissencephalic brains of rodents, the IC is exposed on the lateral surface of the hemisphere, above the rhinal fissure [36] (**Figure 2,** middle). The rodent IC consists of several

heterogeneous subregions that are defined by their cyto-architectural features as well as by their connectivity to other brain regions [37]. In the following paragraphs, I will describe the current state of knowledge about the cyto-architecture as well as the connectivity of the mouse IC.

1.2.1.1 Cyto-architecture of the insular cortex

Across all species, the IC and can be subdivided along the dorso-ventral axis into three major subdivions that are based on their cyto-architecure: the granular (GI), dysgranular (DI) and agranular insula (AI), with the term "granularity" referring to the presence or absence of an internal cortical layer IV, respectively [32,38] (**Figure 2**, right). Thus, while the GI is a classic six-layered cortex, there is a gradual loss of granular layer IV along the DI until in the AI, layer IV is lacking entirely.

The IC can be further subdivided along its rostro-caudal axis into an anterior and posterior part (aIC and pIC, respectively), but these sub-regions cannot be clearly delineated according to their cyto-architecture. This segmentation is rather based on differences in connectivity and function [35]. The insular subregions are strongly interconnected, along the dorso-ventral, as well as along the anterior-posterior axis [39–41]

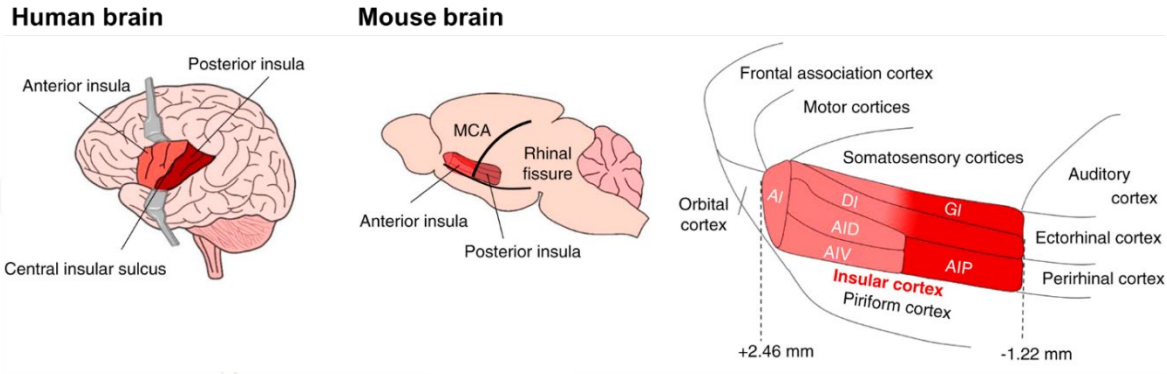


Figure 2: Location and subdivisions of the insular cortex in the human and mouse brain.
Left: In the human brain, the IC is located in the depth of the Sylvian fissure. Middle: In the lissencephalic brains of rodents, the IC is exposed on the lateral surface of the hemisphere above the rhinal fissure. The IC in general can be subdivided along its rostro-caudal axis into an anterior and posterior part. Right: Along its dorso-ventral axis, the IC can be divided into three major subdivisions that are based on their cyto-architecure: the granular (GI), dysgranular (DI) and agranular insula (AI). AID *agranular insula dorsal part,* AIV *agranular insula ventral part,* AIP *agranular insula posterior part. Image adapted from* [42].

1.2.1.2 Connectivity of the mouse insular cortex

The IC is known to be a major, multi-modal integration hub as it is heavily interconnected with different cortical (including the IC itself) and subcortical brain regions. Indeed, in an extensive anatomical tracing study where I was a co-author, we have recently mapped the IC's input and output patterns [41], highlighting the intricate intra-insular as well as brain-wide connectivity of the insular subregions. This complex connectivity pattern gives rise to the IC's multi-modal function in sensory, emotional, and cognitive processes.

1.2.1.3 Internal and external sensory inputs

The insula receives strong sensory information from almost all external senses (e.g. including auditory, olfactory, gustatory and somatosensory signals) via thalamic and cortical inputs [32,37,40,43]. Importantly, the insula, in particular the posterior portion, is the main cortical region receiving information from internal senses, also called 'visceral' or 'interoceptive' sensory inputs (including signals from the intestinal tract, heart, lungs, stomach or bladder) [23,27]. These interoceptive signals travel via the vagus and glossopharyngeal nerves and synapse at the nucleus of the solitary tract (NTS) [44] which in turn connects to the parabrachial nucleus (PBN). The PBN then projects either indirectly via the ventral posteromedial nucleus (VPM) or directly to the posterior insula [40,45–47].

Some of these projections target specific hot spots in the insula, resulting in a topographical organization of the sensory part of the IC, such as for example the 'gustatory', 'visceral' or 'auditory' IC [34,48–50]. It is of specific importance to mention that despite these sensory hot spots, these regions not only process their major sensory input, but they all receive strong multi-sensory afferents. Therefore, the IC is often referred to as a 'multisensory integration hub' [21,48,49].

1.2.1.4 Limbic and cortical connectivity

The IC is heavily interconnected with the limbic system that is responsible for regulating behavioral and emotional responses. The strongest connectivity partners of the IC within the limbic system, but also throughout the whole brain, are the diverse sub-nuclei of the amygdaloid complex, which exhibit diverse and highly specialized connectivity patterns with different subregions of the IC[41]. For instance, our recent study has highlighted a strong, unidirectional projection from the posterior insula to the central nucleus of the amygdala (CeA) and a bidirectional connection between anterior insula and basolateral nucleus of the amygdala (BLA) [35,41]. Other heavily connected limbic areas are for example the mediodorsal nucleus

of the thalamus, the lateral hypothalamus, the bed nucleus of the stria terminalis or the nucleus accumbens [37,41,51]. Interestingly, the IC, in particular the anterior portion, strongly projects to the nucleus accumbens core (NAc) [52,53] which is implicated in regulating positively and negatively motivated behaviors as well as in addiction related behaviors [54–57]. The insula also reciprocally connects to higher cortical areas that are implicated in emotional and cognitive processes, such as the anterior cingulate, orbitofrontal and prefrontal cortices [35,37,38,58,59]. Regarding intra-insula connectivity, it has been suggested that the general information flow in the insular follows a caudal-to-rostral direction [26,60]. Indeed, we recently corroborated this feedforward projections by demonstrating that there are more intra-insular outputs from the pIC than the aIC [41]. Further, the laminar subdivisions of the insula, AI, DI and GI, form strong interconnections along the dorso-ventral axis [32,40].

1.2.1.5 Connections to autonomic centers

The IC not only receives internal bodily information, but also has strong projections to autonomic nuclei. For instance, the insula, in particular the anterior part, has a direct, substantial and topographically organized projection to portions of the NTS that are involved in gustation, general visceral [45] or cardiorespiratory sensations [61]. Moreover the insula also directly projects to brainstem nuclei involved in autonomic motor control, such as the PBN[46], the 'gastric' part of the dorsal vagal complex that is involved in controlling gastric motility via the vagus nerve, or the lateral hypothalamus (LH) that is involved in cardiovascular regulation [45–47,62]. Interestingly, insular neurons innervate both pressor as well as depressor sites (enhancement/inhibition of sympathetic adrenergic activity and elevation/decrease of arterial pressure, respectively) of the LH [62].

Taken together, the IC displays a heterogeneous and extensive connectivity between insular subregions as well as to limbic- sensory-, and executive brain regions, thus placing the insula in a unique position for cross-modal and cross-functional associations [42].

1.2.2 Functions of the insular cortex

The highly complex and wide-ranging connectivity suggests an equally multifaceted functional role of the IC. Indeed, a wealth of human functional MRI studies and electrophysiological recordings in primates and rats have implicated the IC in a plethora of functions, but still, the exact roles of the insula remain elusive. In the following paragraph, I will focus on a few, most coherently described functions that are of relevance to this thesis.

1.2.2.1 Interoception and Autonomic Functions

The IC receives sensory information from the external world, as well as from inside the body - the latter is also called 'interoception'. Interoception is in general defined as 'the sense of the physiological condition of the entire body' [27]. More recently, this definition has been extended to 'the process by which the nervous system senses, interprets, and integrates signals originating from within the body' [63], highlighting a crucial role of an intact interoceptive system in maintaining homeostatic balance and bodily functions and thus ensuring survival. Interoceptive signaling is manifold and includes visceral, cardiovascular, pulmonary, nociceptive, chemosensory, osmotic, thermoregulatory, and autonomic processes [64–70]. The IC, also sometimes referred to as the 'visceral' or 'interoceptive' cortex [71], is one of the main cortical regions involved in interoceptive processing, as human fMRI studies have shown that the IC is strongly activated by the aforementioned interoceptive processes (**Figure 3**) [27].

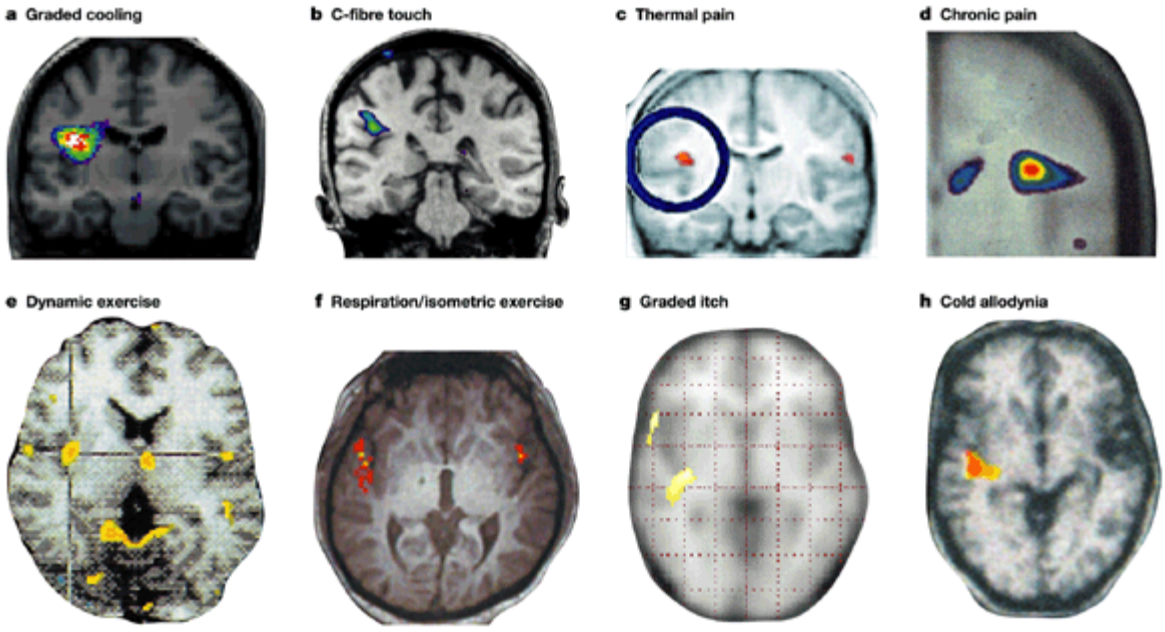


Figure 3: Activation of the interoceptive insular cortex by various modalities.
a) *Graded cooling.* ***b)*** *Sensual (limbic) touch evoked by slow brushing.* ***c)*** *Thermal pain.* ***d)*** *ChronicPain.* ***e)*** *Dynamic exercise on a bicycle.* ***f)*** *Forced respiration and isometric exercise with a hand grip.* ***g)*** *Graded itch elicited by cutaneous histamine injection.* ***h)*** *Cold allodynia. Image adapted from* [27].

Further, it receives topographically organized afferents from distinct thalamic and brainstem relay nuclei involved in the transmission of bodily signals [40,72]. For example, the IC receives input from the NTS,

either directly via the ventral posteromedial nucleus of the thalamus (VPM) or indirectly via the parabrachial nucleus (PBN) [73]. The NTS in turn is the main relay center of the vagus nerve that transmits bodily signals from the peripheral to the central nervous system [74,75]. Importantly, as mentioned above, the IC is not only involved in sensing the body's current condition, it also exerts strong top-down control of autonomic functions, for example regulation of the heartbeat, respiration or bowel movements [42,70,76,77]. Most likely, these functions are mediated through the IC's direct projections to the lateral hypothalamic area, the PBN, and the NTS [45,46]. Interestingly, impaired interoception is considered to be an important element of different psychiatric conditions, such as eating, mood or anxiety disorders or addiction [63].

1.2.2.2 Emotion processing

While the above mentioned studies suggest a vital role for the pIC in processing bodily sensations and homeostatic states, human imaging studies have further suggested an important role for the IC in emotion processing and regulation [24,78]. Indeed, the IC is implicated in processing positive as well as negative emotions, such as anger, sadness, disgust, pleasure, happiness or joy, trust, surprise, social emotions and aversive emotional stimuli [26,42,72,79,80]. Interestingly, the involvement of the insula in emotion processing is cross-modal: the IC is engaged by emotions that are elicited through different modalities (for example language, sounds, pictures or even touch), further highlighting the multi-faceted capacities of insular circuits. In a recent study from our lab that I supported as a co-author, we identified a role for the pIC in aversive state coding and behavioral adaptation to bodily and emotional adversity[35]. Furthermore, we also found that distinct emotion states are encoded at the single cell level within the pIC, highlighting its prominent role in processing various types of emotions[35].

1.2.2.3 Regulation of fear and anxiety

Particularly strong evidence supporting a vital role of the IC in mediating fear and anxiety comes from studies in humans. The IC has been shown to co-activate with a collection of brain regions commonly referred to as 'fear-network', including the amygdala, ventral hippocampus, bed nucleus of the stria terminalis, medial-prefrontal and sensory cortices [81]. For example, the insula is activated while healthy subjects were anticipating and/or seeing aversive images [82]. Further, studies using specific pharmacological agents to provoke fear or anxiety states in healthy individuals (e.g. cholecystokinin-4, procaine or alpha-2 adrenergic antagonists) showed increased activation in insula [83–85]. Also, a stronger connectivity between insula and amygdala is associated with greater severity of trauma-related symptoms in subjects that repeatedly experienced traumatic events [86]. Interestingly, insular activity also correlates

to trait anxiety, as anxiety-prone individuals show an higher insula activation when exposed to aversive stimuli in comparison to anxiety-normal subjects [82,87]. Additionally, experiments using the classical Pavlovian fear-conditioning paradigm have demonstrated specific fear-induced activation of the IC [88,89]. Moreover, the human insula has emerged from numerous meta-analyses as a core region affected in anxiety disorders. For instance, patients suffering from posttraumatic stress disorders (PTSD), generalized anxiety disorders, specific phobias or social anxiety show altered activity in the IC [90–93]. Interestingly, some studies indicate that anxiety disorder correlated hyper-activation in the insula is reduced upon successful therapy [94–96].

There are only a few animals studies that support altered insula activity in fear and anxiety behaviors: electrophysiological recordings or calcium imaging from our own lab in awake behaving animals showed increased insular activity when being exposed to a fear-evoking cue after auditory fear conditioning [97] or when animals explore anxiogenic environments [35]. The insula can also directly influence fear and anxiety behaviors. For instance, optogenetic, chemogenetic or pharmacological insula manipulations influence innate fear behaviors, such as exploration of anxiogenic environments or the exposure to innately aversive stimuli such as predator odor [35,98–100]. Further, evidence from lesion studies and pharmacological inhibition of different insula subregions in the rat have demonstrated an involvement of the insula in the acquisition and/or consolidation of learned fear [101–104]. However, the results of the above-mentioned studies are contradictory, as some suggest that the insula is exclusively involved in acquisition or consolidation of learned fear behavior, while others hypothesize that it is important for the expression of fear behaviors. Further, there is some evidence that discrete subsets of pIC neurons change their activity upon fear conditioning and subsequent extinction [97], but the causal involvement of the pIC in fear extinction has not been studied yet. Interestingly, the insula has also been found to play a role in the learning of safety cues, which inhibit the expression of conditioned fear [105–108], suggesting that fear promoting and inhibiting circuits may co-exist within the insula.

1.2.3 Altered interoception and anxiety

While the mechanistic role of the insula in fear and anxiety remains unclear, there is a steadily growing body of neuroscientific research aiming at understanding the link between anxiety and altered interoception. Interoception is considered to be comprised of two components [25]: on the one hand, interoceptive sensory inputs are often immediately associated with intense affective and motivational components. For example, the sensation of pain when touching a hot stove, even though only experienced in the peripheral tissue, is instantly attributed to the object that evoked the painful sensation and causes a quick withdrawal reflex. On the other hand, the interpretation of those interoceptive signals strongly

depends on the current homeostatic state of the individual. For instance, drinking a mildly salty solution can be either rewarding or punishing, depending on the current osmotic state of an individual[109]. As mentioned earlier, the IC is a site where bodily sensations, autonomic control and afferents from brain regions implicated in emotion processing converge, suggesting that it lies exactly at this interface between bodily and affective states. The insula creates an image of an individual's current bodily and affective state by integrating those internal and external senses. Indeed, human and rodent imaging studies have implicated the insula in the representation of physiological states such as hunger or thirst [110–114]. Additionally, in a study where I was a co-author, our lab has recently demonstrated, that neurons in the mouse IC encode long-lasting states of anxiety [35].

Moreover, there is strong evidence that the insula is critically involved in predictions about future bodily states by computing 'interoceptive predictions' [25,115–118]. That means it generates anticipatory signals about changes in the physiological state that are important for learning about appetitive or aversive outcomes. Consistent with this hypothesis, human and rodent insula responds to predictive cues across different sensory modalities [110,111,119,120]. But how is impaired interoception involved in the generation of anxiety disorders?

In 2006, Paulus and Stein have proposed that an altered signal about an impending aversive bodily state constitutes the basic link between interoception and anxiety [25,121]. They showed that individuals who are prone to anxiety are more aware or focused on their bodily signals, thus generating increased interoceptive prediction signals about future aversive physical states. This in turn may therefore trigger anxiety, worry and avoidance behaviors. Experimental evidence for this theory comes from several studies showing that anxiety disorders, including panic disorder, PTSD and social phobia in particular, were found to be closely correlated to the level of interoceptive awareness [122–128]. Strikingly, interoceptive awareness is also positively related to the activity in the IC [129–131].

In summary, given its unique role in mediating fear as well as interoception, the insula represents a promising avenue to study and understand the precise neuronal mechanisms underlying pathological fear and anxiety.

1.3 Experimental approaches to study the role of the insula in learned fear behaviors

1.3.1 Behavioral experiments in mice

The identification of neural circuits underlying particular behaviors is the first step to understand how the brain generates this behavior [132]. To uncover the relation between brain activity, sensory processing, cognition and behavioral outcome, the most commonly used method in human neuroscientific research is functional magnetic resonance imaging (fMRI) that measures brain activity indirectly via blood flow while subjects are exposed to sensory stimuli and/or engaged in a behavioral task. However, this approach is limited regarding temporal and spatial resolution, and the correct interpretation of the data remains a major challenge [133]. Additionally, intervention studies in humans to investigate the functional involvement of a brain area in certain behaviors or processes are restricted to subjects suffering from brain lesions or patients with implanted deep brain stimulation electrodes (e.g. as used for the treatment of Parkinson's disease [134]). Thus, in the last decades, neuroscience has increasingly turned towards model organisms to study how the brain integrates sensory processes to generate certain behaviors. Specifically rodents are among the most commonly used animal models in neuroscientific research. They are well-suited model organisms, as they display a variety of behaviors that are relevant for human behaviors and diseases [135]. Additionally, over last 2 decades a large toolset of activity manipulations, large-scale recording techniques, genetic analyses and modifications as well as precise anatomical dissection tools have been developed to study the nervous system in unprecedented detail in mice as animal models. Thus, mouse models enable the detailed analysis of the influence of certain genes, cell types or brain regions on physiological and pathophysiological brain functions and behavior. However, it is important to bear in mind that almost all commonly used rodent behavioral tests model only certain domains of a behavior, and trying to model an entire human pathology might be an unrealistic attempt [136]. Therefore, most rodent behavioral tests use a reductionist approach to make a very specific set of behaviors applicable to neuroscientific analysis. Hence, rodent behavioral experiments represent a simplified, yet powerful tool to eventually tackle the underlying causes for maladaptive disorders, such as anxiety, depression, or addiction.

A rich repertoire of behavioral assays has been established as a standard in rodents to test diverse brain functions, from basic locomotor and sensory functions, to more complex behaviors such as learning, cognition and emotionality[135]. In particular, a large variety of widely employed tests is commonly

employed to study fear and anxiety related behaviors in rodents [20,135,137–139]. On the one hand, some of those experiments test for an animal's innate fear-related behaviors, such as the *Open Field Test* (OFT) or the *Elevated Plus Maze* (EPM). Both setups exploit a mouse's different unconditioned and internal drives: fear of open spaces/heights (avoidance), tendency toward dark, enclosed spaces as well as exploration driven behaviors (approach) [140,141]. On the other hand, and most important for the present study, learned fear behaviors and their inhibition can be effectively modeled using a classical fear condition and extinction paradigm (as already described in detail in 1.1.2). Fear conditioning and extinction have first been described by Pavlov in 1927 [12] and have since been an invaluable tool for neuroscientists, not only to study basic neuronal mechanisms of a relatively simple learning process, but also helping in understanding the neuronal underpinnings of acquired fear and anxiety behaviors (see [19,142,143] for comprehensive reviews). In particular, fear extinction experiments have become particularly appealing for translational research for several reasons: Firstly, the readouts are easy to compare between humans and rodents, given the similarity of laboratory testing procedures [139]. Secondly, many of the brain regions that have already been identified to be involved in fear extinction are common across humans and rodents, supporting the idea that observations made in one species are informative to the other [144]. Thirdly, impaired fear extinction has been linked to a variety of psychiatric conditions such as PTSD, phobias, panic or obsessive-compulsive disorders [145–147]. Lastly, the principle of fear extinction, where an unwanted fear response is diminished by repeated exposure to the fear evoking stimulus, is a core process underlying exposure therapy [148]. Taken together, fear conditioning and extinction experiments in mice represent a simple, yet powerful tool for studying the neuronal underpinnings of learned fear behaviors as well as the inhibition of those behaviors in the laboratory.

1.3.2 Optogenetics

'Optogenetics' is the combination of genetic and optical tools to cause an activity increase or decrease in targeted cells in a wide range of biological systems (from tissue cultures to insects, rodents and even primates) *in vitro* as well as *in vivo* [149]. The development of modern optogenetic tools allowed the precise manipulations of neuronal activity with a high temporal and spatial precision [150,151]. The general principle of optogenetics is to exploit the light-sensitivity of microbial opsins to regulate the flow of electrically charged ions across neuronal membranes. Opsins are seven-transmembrane protein channels that, upon illumination with the appropriate light wavelength and pattern, lead to either activation or inactivation of neurons with millisecond precision (**Figure 4**). Today, a large palette of opsins has been identified and optimized regarding light sensitivity, kinetics, ion-specificity, and expression profiles [152–154].

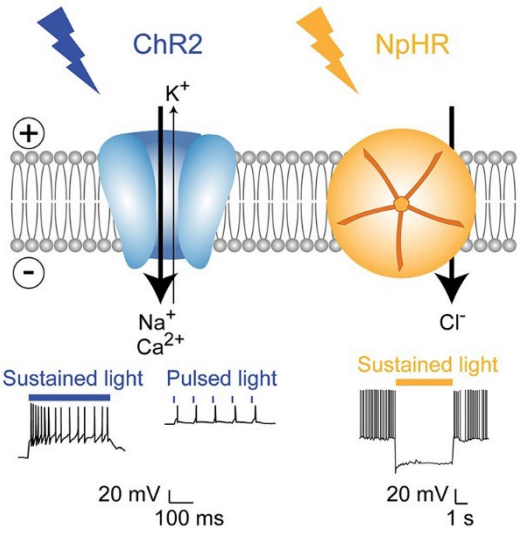


Figure 4: Optogenetic approach for neuronal activation using Channelrhodopsin (ChR2) and inactivation using Halorhodopsin (NpHR).
Upon illumination with the appropriate wavelength, the ion channels open, leading to an influx of Na^+ or Ca^{2+} (in the case of ChR2) or Cl^- (in case of NpHR) into the cell, thus either depolarizing or hyperpolarizing the neuron, respectively. Whereas for ChR2, pulsed light with specific frequencies is necessary to activate the neuron, cells are inactivated with NpHR using sustained light. Image adapted from [155].

The *in vivo* use of optogenetics has led to unprecedented advances and has become an indispensable tool in modern neuroscientific research. From innumerable studies using optogenetic manipulations, we have gained mechanistic understandings of a large number of neuronal circuits and their implications in behavior. However, an extensive introduction of current advances in optogenetics is not within the scope of this thesis and hence I want to refer to many excellent reviews and protocols [150,154,156].

Despite all of this understandable enthusiasm regarding optogenetics, there are caveats when applying optogenetic approaches to study neural circuits *in vivo*. First, when optogenetic activators are used in gain of function experiments, strong stimulation risks driving neuronal responses outside the physiological range. This may in turn cause unnatural plasticity in the circuit, or even engage (or disengage) downstream brain areas that would not be targeted under physiological conditions. Especially when assessing the behavioral outcome of those artificial stimulations, one needs to consider that the evoked behaviors might not occur naturally [132], potentially leading to physiologically incorrect conclusions about circuit function [157]. Additionally, it has been reported that also neuronal deactivators can drive the neuronal activity below their normal range, and the subsequent release of optogenetic silencing may cause a rebound excitation [152,158]. Further, conventional optogenetic tools usually stimulate/inhibit all neurons within a broad, genetically defined population and cannot selectively target specific subtypes within this population, thus making it difficult to disentangle the contribution of neuronal subpopulation to certain behaviors. Therefore, limitations of this approach need to be taken into account when designing, performing and

interpreting optogenetic experiments. Nevertheless, in comparison to traditional tools for neuronal activity manipulations, such as electrical micro-stimulations or pharmacological interventions, optogenetics is still among the best-suited tools to control neuronal activity in a temporally, spatially and genetically defined manner.

1.3.3 Fiber Photometry

The functional characterization of neuronal circuits in rodents requires tools to measure neuronal activity in awake, behaving animals. Neuronal activity causes rapid changes in intracellular free calcium and, these calcium dynamics can thus be used as an approximate readout for neuronal activity. Calcium imaging experiments have relied on this principle to monitor the activity of large neuronal populations [159,160]. The development of genetically encoded calcium indicators (GECIs), for example the GCaMP proteins family [161], opened up a new avenue for cell type-specific and long-term monitoring of neuronal activity *in vivo*. However, two major caveats of calcium indicators are 1) their relatively slow temporal dynamics and therefore their inability to report real-time neuronal activity, and 2) the fact that neuronal activity is only indirectly measured via calcium dynamics, therefore being only a rough proxy for actual neuronal activity. Nevertheless, calcium imaging is being used in combination with manifold approaches, such as two-photon imaging in head-fixed mice [162,163], where the activity of up to hundreds of neurons can be monitored, or single-photon or even multi-photon imaging using head-mounted miniature endomicroscopes in freely moving mice [164]. However, those approaches have their limitations, as the required microscopes are very costly and/or heavy, and thus are not suitable for monitoring neuronal activity in freely behaving mice, respectively. Alternatively, fiber photometry (FP) provides a comparatively easy solution to record bulk neuronal activity of genetically-defined neuronal populations [165,166]. For FP recordings, an optical fiber is chronically implanted near the targeted brain region of interest expressing a GECI. The same optic fiber delivers excitation light and collects overall fluorescence induced by calcium-activity during light excitation. Thus, FP does not monitor single cell activity, but reports the overall fluorescence changes in all calcium indicator-expressing neurons within detection-range of the fiber. A huge advantage of FP (aside of its relatively easy handling) is that the mobility of the mouse is only mildly affected and thus, its behavior and the neuronal correlates can be studied in a more natural, undisturbed manner.

1.3.4 *In vivo* single unit electrophysiology

Since FP recordings suffer from certain limitations such as a relatively poor temporal and non-existing cellular resolution, as well as only indirect neuronal activity measurements, the more classical approach of intracranial electrophysiological recordings still represents a gold standard for understanding precise neuronal dynamics in awake behaving animals. Using either static electrodes, microdrives or more recently also high-density silicone probes [167], it is possible to monitor neuronal activity at various spatial scales, from single cells to whole populations across different brain regions at an extremely high temporal resolution [168–171].

In particular, "*in vivo* single unit electrophysiology" describes a method to extracellularly record the activity from individual neurons in the intact nervous system, usually using microelectrode arrays [172]. The action potential, or "spike", recorded with extracellular electrodes is produced by currents that are induced to flow in the extracellular space around an active neuron. Electrode bundles with low impedance can detect those changes in electrical current as spikes, and, depending on where an electrode tip was positioned along the neuron (soma or axon), these spikes are recorded as a specific waveform with a characteristic spatiotemporal profile [173]. The waveform provides information about the exact timing of a neuronal spiking event, but also about the cell type (interneurons or pyramidal neuron) the signal was recorded from (even though this is not fully reliable as also other features such as firing rate and response dynamics have to be taken into account) [174–176]. However, as multi electrode arrays detect diverse extracellular signals, one major challenge in this approach is the identification and isolation of recorded neurons, and this "spike sorting" procedure can require time and effort. Most commonly, spike sorting is achieved using semi-automated processes that require human post-processing control [177]. To this end, the first step is to detect the spikes, which typically happens by high-pass filtering and thresholding the wide band signal. Second, each waveform is characterized by a principal component analysis into a "feature vector". Third, these vectors are divided into groups corresponding to putative "single units" or "multi unit"-related background signal (noise) using a cluster analysis. Finally, the results need to be manually confirmed, to adjust any errors made by automatic algorithms. In the last decade, several novel approaches for spike sorting were developed that promise highly reliable, fully-automated sorting algorithms [177–185].

Importantly, electrophysiological single unit recordings also only mildly affect an animal's mobility, and thus enabling us to investigate behaviors that involve more natural movements and offering ethologically valid insight into neural activity. However, one caveat of single unit recordings is that it can detect the signals only from a few, randomly distributed neurons located near the electrode tip, thus merely giving an incomplete picture of the larger-scale activity of a certain brain area.

In conclusion, the combination of single unit and FP recordings represents an ideal approach to reveal the global reactivity of larger, spatially related neuronal ensembles.

1.3.5 Vagus nerve stimulation

The vagus nerve is the tenth and longest cranial nerve and is largely composed of 80% afferent and only 20% efferent fibers [186]. However, its cholinergic efferent fibers are the main mediator of the parasympathetic branch of the mammalian autonomic nervous system. Thus, it provides information about peripheral arousal to the brain while sending feedback to activate the parasympathetic nervous system and promote adaptive physiological and behavioral strategies [187,188]. Further, the vagus contains all three main fiber types: A-, B-, and C-fibers, which are distinguished by fiber diameter, myelination, and activation thresholds [74], further highlighting its complexity. The vagus nerve is a bilateral structure, and left and right vagus originate from the brainstem, from where they project through the neck (along the carotid artery), the upper and lower chest, the diaphragm, all the way down into the abdomen (**Figure 5)** [189]. During this course, the vagus heavily branches out and innervates various organs such as the larynx, pharynx, heart, lungs, and gastrointestinal tract [74,190]. As already described above, the primary central target of the vagal sensory fibers is the NTS, that in turn projects to different brain regions, such as the hypothalamus, locus coeruleus (LC), amygdala and, via the PBN and VPM also the viscerosensory portion of the IC [74].

Vagus nerve stimulation (VNS) has its beginnings already in the late 19th century, where manual VNS (via compression of the carotid artery) was described as a treatment of epileptic seizures by means of lowering the heart rate [191,192]. With the development of electric VNS, studies were beginning to investigate the effect of VNS on the central nervous system, showing that VNS causes changes in cortical and subcortical activity [193,194]. In the mid 1980s, the therapeutic benefits of VNS were proven for the first time by demonstrating that VNS can terminate epileptic motor seizures in dogs [195–197] and VNS has first been applied in patients suffering from epilepsy in the early 1990s by implantation of an electric stimulator [198]. Interestingly, some of the following studies reported that patients receiving VNS for epilepsy treatment also showed an improvement in mood and cognition, even patients that did not show an improvement in seizure control [199,200]. Further, it has been reported that VNS affects the metabolism (and therefore the function) of important limbic structures such as the amygdala or the IC [201]. This has led to the application of VNS implants specifically for patients suffering from treatment-resistant depressions [202]. Additionally, more recent studies even report that VNS has beneficial effects on rapid cycling bipolar disorders and treatment-resistant anxiety disorders [203–205].

Very recently, studies using VNS in rodents have indicated that VNS has beneficial effects on innate anxiety behaviors [206,207]. Further, a number of studies suggest that VNS promotes extinction learning and extinction memory consolidation in healthy rats as well as in rat models for PTSD [208–213]. On the other hand, it has been shown that subdiaphragamactic vagal deafferentiation (SDA), where central signaling of the vagus nerve was completely disrupted, also caused a reduction in innate anxiety-like behaviors, but increased expression of auditory-cued fear conditioning and impaired fear extinction [214], leaving the effects of VNS on fear and anxiety behaviors in rodents contradictory and elusive. Additionally, the neuronal circuits mediating the effects of VNS on fear and anxiety behaviors have not been studied yet and remain completely unknown.

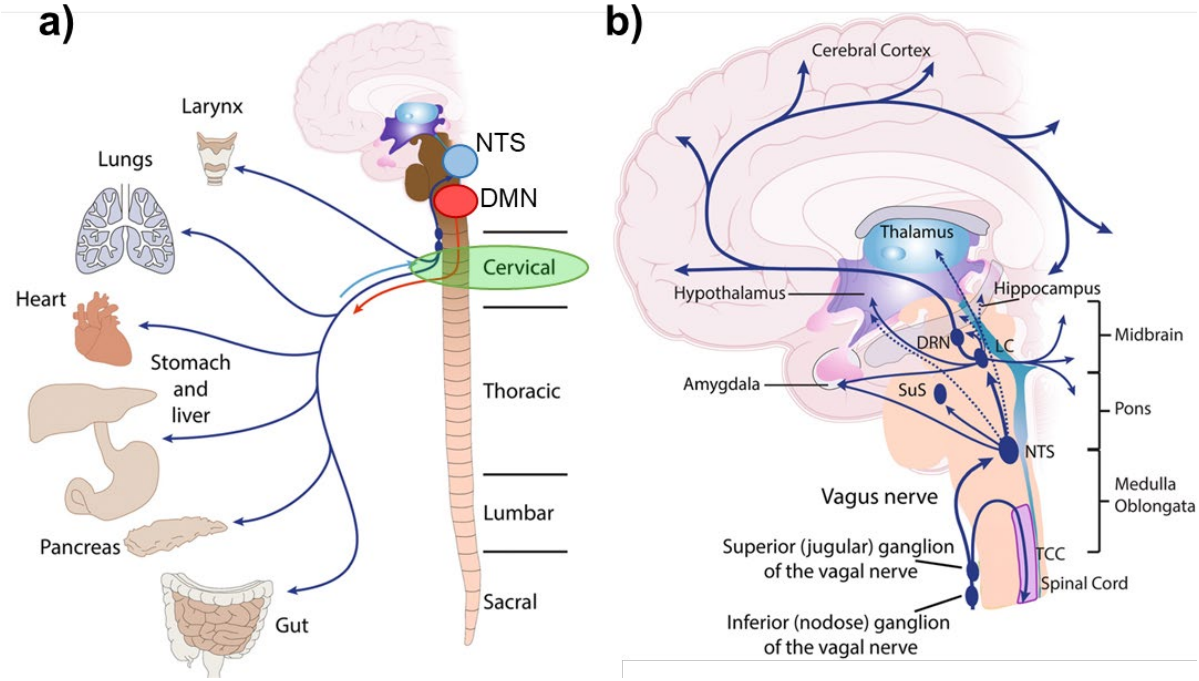


Figure 5: Overview of vagal circuitry linking the central and peripheral nervous system.
a) *Vagal efferents mostly emerge from the dorsal motor nucleus (DMN) and are key to exerting autonomic control in the periphery.* ***b)*** *Vagal afferents from the superior and inferior vagal ganglia mostly converge onto the nucleus of the solitary tract (NTS) in the brain steam, from where it projects directly or indirectly to a large variety of other brain regions. For vagus nerve stimulation (VNS), the cuff is placed around afferent and efferent fibers of the vagus nerve at cervical level (green circle in* ***a)****).*
DRN *dorsal raphe nucleus,* LC *locus coeruleus,* SuS *superior salivatory nucleus (preganglionic parasympathetic neurons),* TCC *trigeminocervical complex. Image adapted from* [215].

1.4 Aims of the study

Studies in humans as well as rodents have suggested that the IC plays an important role in the regulation of fear and anxiety and, in particular, in the acquisition and consolidation of conditioned fear behaviors. However, the involvement of the pIC in fear extinction remains elusive.

Therefore, **the first goal** of my thesis was to characterize fear related neuronal activity in the mouse posterior IC when animals were undergoing a classical fear and extinction behavior experiment. In order to first broadly characterize the dynamics of the pIC during fear learning and extinction, I applied fiber photometry in freely moving mice to record from bulk neuronal activity in the pIC. Second, to gain deeper insights into the dynamics of individual neurons during fear conditioning and extinction, I also applied single unit electrophysiological recordings to monitor pIC activity at single cell level in freely behaving mice.

The second goal was to characterize the potential involvement of the pIC in fear extinction. Towards this goal, I employed optogenetic manipulations in freely moving mice to inhibit the pIC during fear extinction. Further, as the initial results from optogenetic experiments were paradoxical, I aimed to disentangle the influence of an animal's internal fear state on the effects of optogenetic pIC inhibition during fear extinction.

Based on the results that I obtained from the second goal, **the third goal** of my thesis was to gain insight into how the internal fear state can influence pIC activity and fear extinction performance. I hypothesized that the internal fear state has a strong bodily component that might be transmitted to the brain via the vagus nerve. To study the influence of the vagus nerve on extinction learning, I applied vagus nerve stimulation (VNS) when animals were undergoing fear extinction while monitoring pIC activity using fiber photometry.

2 MATERIALS AND METHODS

2.1 Animals

Animal experiments were conducted in accordance with the regulations from the government of Upper Bavaria and by the committee on Animal Health and Care of the *Institut National de la Santé et de la Recherche Médicale* and French Ministry of Agriculture and Forestry. Male C57BL/6NRj mice were received from the in-house breeding facility (Max Planck Institute of Neurobiology / Biochemistry, Martinsried). All mice that underwent behavioral testing were between the age of 2–6 months, housed in pairs of two and kept on an inversed 12 h light/dark cycle (lights off at 11:00 am). Mice were provided with *ad libitum* access to standard chow and water.

2.2 Viral constructs

For *in vivo* optogenetic experiments, the following viral constructs were obtained from the UNC Vector Core (Gene Therapy Center, University of North Carolina at Chapel Hill, USA): AAV2/5-Camk2a-eNpHR3.0-eYFP (5.2×10^{12} particles per ml), AAV2/5-CamKIIa-hChR2(H134R)-eYFP (6.2×10^{12} particles per ml) or AAV2/5-CamKIIa-eYFP (4.3×10^{12} viral genomes (vg) ml^{-1}) as control virus. For fiber photometry experiments, I used AAV9.Camk2a.GCaMP6s.WPRE.SV40 ($\geq 1 \times 10^{13}$ vg ml^{-1}) obtained from Addgene (viral preparation #107790-AAV9). All adeno-associated virus vectors were obtained under MTAs with the University of North Carolina.

2.3 Surgeries

2.3.1 Stereotactic Surgeries

30 minutes before surgery, Metamizol (Hexal, 200 mg/kg bodyweight, s.c.) was applied for peri-operative analgesia. Anaesthesia was initiated with 5% isoflurane in oxygen-enriched air and maintained at 1-2.5% throughout the surgical procedure. Mice were firmly secured in a stereotaxic frame (Model 51733D, Stoelting) and placed on a heating pad to maintain their body temperature at 37 °C. Eye ointment (Bepanthen, Bayer) was applied to protect the animal's eyes. For viral injections, glass-pipettes (B100-50-10, Sutter Instruments) were pulled and attached to a microliter syringe (5 µL Model 75 RN, Hamilton) using glass needle compression fittings (#55750-01, Hamilton). The syringe was mounted onto a syringe pump (UMP3, WPI) that was attached to the stereotaxic frame and controlled by a microcontroller (micro4, WPI). Before skull trepanations, Lidocaine was applied topically at the drilling site. Coordinates used to target the pIC were at the following distances from Bregma: AP: -0.40 mm, ML: ± 4.10 mm, DV: - 4.00 mm.

For optogenetic experiments, two skull trepanations were performed above the pIC. Then, 250 - 300 nl of optogenetic virus were bilaterally injected at a rate of 50 - 100 nl/min into the pIC at DV - 4.00 mm. Custom made optic fibers fibers were tested before implantation to transmit ≥80% of the incoming light. Then, these fibers (200 µm core diameter, 0.22 NA, in 1.25 mm zirconia ferrules, Thorlabs) were secured with acrylic glue (Ultra Gel, Pattex) 0.50 mm above the injection site. Additionally, I added a layer of black dental cement (Super Bond C&B, Génerique International) to further secure the optic fibers and to reduce light emission from the skull.

For fiber photometry experiments, a single skull trepanation was performed above the left pIC and 200 - 300 nl of photometry virus (AAV9.*Camk2a*.GCaMP6s.WPRE.SV40) were unilaterally injected at a rate of 50 - 100 nl/min at DV - 4.00 mm. Custom made optic fibers (200 µm core diameter, 0.48 NA, in 1.25 mm zirconia ferrules, Thorlabs) were tested before implantation to transmit ≥80% of the incoming light. The fiber was secured using acrylic glue (Ultra Gel, Pattex) and black dental cement (Super Bond C&B, Génerique International) 0.2 - 0.3 mm above the injection site.

For in vivo electrophysiology experiments that were conducted in our collaborator's laboratory in Bordeaux, a custom made electrode bundle consisting of a 16-wire electrode array (Nichrome wires, 13 µm in diameter, impedance of 30–100 KΩ; Kanthal) attached to an 18-pin connector (Omnetics) was used. The 16 electrode wires were cut at an angle such that all wires extended to slightly different depths below the optic fiber. A single skull trepanation was performed above the pIC and the electrode bundle

was slowly lowered into the left or right pIC at DV - 4.00 mm and secured using acrylic glue (Ultra Gel, Pattex) and black dental cement (Super Bond C&B, Générique International). Additionally, optrodes were referenced/grounded via a silver wire (127 µm in diameter; A-M Systems) placed above the cerebellum.
For in vivo *optrode experiments,* the animals were treated as described above, with the exception that they first received an injection of 300 nl of optogenetic virus (AAV2/5-*Camk2a*-eNpHR3.0-eYFP or AAV2/5-*Camk2a*-hChR2(H134R)-eYFP) into the pIC. Additionally, a custom made "optrode" that consisted of an electrode bundle as described above combined with an optic fiber (200 µm diameter, 0.22 NA) was implanted. The electrodes were cut at an angle so that the longest electrode wire was extending 0.5 mm below the optic fiber, so that the fiber was placed 0.5 mm above the virus injection site.
At the end of each surgery, animals were closely monitored while waking up from anesthesia and Carprofen (Rimadyl, Pfizer, 5 mg/kg bodyweight, s.c.) was applied for post-operative pain care. Experiments started between two (electrophysiology) and five weeks (optogenetics and fiber photometry) after the surgery to ensure full recovery from surgery and sufficient virus expression, respectively.

2.3.2 Surgeries for Vagus Nerve Stimulation

For vagus nerve stimulation (VNS) experiments, animals were anesthetized with a mixture of Midazolam, Medetomidin and Fentanyl (5 mg/ kg, 0.5 mg/kg and 0.05 mg/kg bodyweight, respectively, i.p.). The animal was placed on a heating pad to maintain its body temperature at 37°C and eye ointment (Bepanthen, Bayer) was applied to protect the animal's eyes. The left vagus nerve was accessed at the cervical level through an incision in the skin along the ventral midline. The muscle layers were separated, exposing the left vagus nerve and carotid artery. Then, a custom-made vagus nerve cuff consisting of platinum- iridium wire electrodes and biocompatible micro-renathane cuffs (0.04 in. i.d., 0.08 in. o.d., 4 mm long) attached to a 2-pole connector (Omnetics) was wrapped around the left vagus nerve. Transient cessation of breathing was observed in anesthetized rats when VNS (0.8 mA, 60 Hz, 50ms for 5 sec) was given through the implanted electrode, confirming that the cuff electrode was appropriately positioned to stimulate the vagus nerve. The connector was then tunneled under the skin and fixed to the skull using acrylic glue and dental cement. In experiments where VNS was combined with photometry recordings, 300 nl photometry virus (AAV9.*Camk2a*.GCaMP6s.WPRE.SV40) was injected into the left pIC and a custom made optic fiber (200 µm core diameter, 0.48 NA, in 1.25 mm zirconia ferrules, Thorlabs) was implanted 0.3 - 0.7 mm above the injections site as described above.
At the end of the surgery, the anesthesia was reversed with a mixture of Naloxone, Flumazenil and Atipamezole (1.2 mg/kg, 0.5 mg/kg, and 2.5 mg/kg bodyweight, respectively, i.p.) and the animal was

closely monitored while waking up from anesthesia. Carprofen (Rimadyl, Pfizer, 5 mg/kg bodyweight, s.c.) was applied for post-operative pain care.

2.4 Behavior experiments

All mice were handled daily by the experimenter for around 5 minutes for at least five days before the start of behavior experiments. Additionally, for optogenetic, fiber photometry and *in vivo* electrophysiology experiments, animals were habituated to being tethered during the handling procedure. All behavioral experiments were performed during the dark phase of the light cycle between 12:00 and 19:00.

2.4.1 Fear conditioning and extinction

Mice were subjected to an auditory fear conditioning and extinction (FC/Ext) paradigm by using a commercial fear conditioning setup (Ugo Basile, Italy), consisting of a behavior box, an electric grid floor, a light and a sound source (Tucker Davis Technologies) located in a sound proof chamber. Two different contexts were used depending on the stage of experiment (context A and B) that had different olfactory, visual and haptic features. In context A, the behavior box and the grid floor were cleaned before and after each session with 80% ethanol, the light intensity was set at 250 lux and the walls were covered with a distinct visual cue. In context B, the cleaning agent was 1% acetic acid, light intensity was at 200 lux, the floor was covered with plastic and the walls of the box were transparent. As a readout of an animal's fear response, I used freezing behavior (defined as complete immobility with the exception of breathing) which is a characteristic fear response in rodents. Freezing was scored with the software ANYmaze (Stoelting): Behavior of the mice was videotaped and the software calculated a "freezing score" depending on the number of pixel changes between frames. If the freezing score fell below a certain threshold for at least 2 seconds, mice were considered to be freezing. In some animals, freezing behavior during extinction training was manually corrected *post hoc* to exclude that immobility was incorrectly detected as freezing behavior.

The standard FC/Ext protocol consisted of 4 days: On day 1, mice were habituated in the morning to four presentations of the later conditioned stimulus (CS+; pips of 7.5 kHz, 250 ms, repeated at 2 Hz for 30 s) alternating with a control tone (CS–; white noise for 30 s) with random ISIs (ranging between 30 and 130 s) over a period of 15 min. Fear conditioning (FC) was performed in the afternoon of the same day by pairing five CS+ presentations with a foot shock (unconditioned stimulus, US). The US strength was 0.4 mA and 1 s in duration. The CS– tone was presented after each CS+–US association but was never

reinforced, the ISIs ranged between 30 and 120 seconds and the total experiment duration was 20 minutes. On day 2, fear recall was tested in same context A where the animals were first exposed to the context only for 300 seconds to test for contextual fear memory. Afterwards, four CS+ and four CS– were presented in a randomized order with random ISIs ranging between 30 and 60 seconds (total experiment duration of 16 min). On day 3, fear extinction training took place in Context B, where the CS+ was presented twelve times with random ISIs ranging between 30 and 90 seconds (total experiment duration of 20 minutes). On day 4, fear extinction memory was tested in the same context B by presentation of two CS+ with an ISI of 40 seconds and a total experiment duration of 4 minutes.

In the case of 'consecutive weak and strong FC/Ext', three days after the first FC/Ext protocol that was conducted as described above, animals were undergoing another round of FC/Ext. This paradigm is a shortened version of the previously described FC/Ext with the following execeptions: day 1 was started immediately with FC in a novel context C (100% Isopropanol, distinct visual cues on the walls, light intensity at 200 lux). A novel conditioned stimulus (CS+2, upsweeps from 5 to 15 kHz, 500 ms duration, repeated at 1 Hz for 30 s) was paired five times with a "strong" US (0.6 mA and 2 s in duration), whereas the same CS- (white noise, 30 s) was presented alternatingly five times but never reinforced. On day 2, animals were immediately undergoing extinction training in a novel context D (soap smell, 20 lux, different visual cues on the walls) where they were exposed to twelve CS+2 presentations. Extinction memory was tested in an extinction retrieval on day 3 where animals were exposed to two CS+2 in the same context D.

In the case of 'shortened FC/Ext', the protocol was similar to the standard FC protocol, with the following exceptions: the recall consisted of only 2 CS+ and 2 CS- presentations, and during extinction, only 8 CS+ were presented (note that here, the US had an intensity of 0.4 mA and was 1 s long ("weak FC")).

For all FC/Ext paradigms, I analyzed freezing behavior by calculating the % time an animal spent freezing during a CS presentation, and pooled CS evoked freezing across two or four consecutive CS presentations for each animal (the respective pooling method is mentioned in each figure legend).

2.4.2 Optogenetic inhibition during extinction training

The animals were undergoing habituation, fear conditioning and recall without any interventions as described above. On the day of extinction, both implants were tethered to a yellow laser source (MGL-F, 593.5 nm wavelength, CNI Lasers) via fiber-optic patch cables and a rotary joint (0.22 NA, Doric Lenses, Canada) allowing the animal to freely move around in the behavior boxes. The laser intensity during all experiments was set to 12 - 20 mW at the tip of the patch cables and the laser was synchronized via TTL

(transistor-to-transistor logic) pulses to ANYmaze software. For pIC inhibition during extinction training, constant light was delivered during each CS+ presentation as well as 5 seconds before and after the tone (40 s inhibition in total).

2.5 Fiber photometry recordings

2.5.1 Photometric signal acquisition

As described previously[216], I used GCaMP6s bulk fluorescence measurements through a single optical fiber for both delivery of excitation light and collection of emitted fluorescence. For both excitation and emission measurements, I used a commercial fiber photometry system (one site, two color; Doric Lenses) with two excitation wavelengths, 405 nm (isobestic point of the GCaMP signal; calcium independent) and 465 nm (GCaMP signal; calcium dependent). The two excitation LEDs were fiber coupled into patch cords. Excitation light intensity was measured at the end of these patch cords as ~20 µW for 465-nm light and ~7 µW for 405-nm light. On the day of the experiment, the fiber-optic implant was coupled to the fiber photometry system via low-autofluorescence fiber-optic patch cables (0.48 NA) and a rotary joint (all from Doric Lenses, Canada). The fiber photometry system was synchronized with ANYmaze software using TTL (transistor-to-transistor logic) pulses. Photometry signals were recorded at 12 kHz, demodulated and downsampled to 30 Hz for analysis. pIC activity was monitored throughout the whole FC/Ext procedure.

2.5.2 Photometric signal analysis

All photometry data were analyzed with custom-written MATLAB programs. To remove the strongest photobleaching artifacts, I first excluded the first 10 s of each recording and the data were smoothed using a second order Savitzky-Golay filter with a frame length of 21 frames. To remove bleaching and motion artifacts, a least-squares linear fit was applied to the 405-nm control signal and aligned to the 465-nm signal, by using a procedure developed by Lerner et al.[216] (https://github.com/talialerner/Photometry-Analysis-Shared/blob/master/Dropbox/MATLAB/Shared%20photometry%20code/controlFit.m). The change in fluorescence (ΔF/F) was calculated as ΔF/F = (465-nm signal – fitted 405-nm signal)/fitted 405-nm signal. Event related changes in calcium signals were calculated by z-scoring the ΔF/F values. To calculate CS evoked response, the z-score was built by using the mean and SD from the 30 s before tone onset (“preCS”); for US evoked responses, the z-score was built by using the mean and SD from the 1 s

before US onset; for freezing evoked responses, the z-score was built by using the mean and SD from -6 s to -4 s (2 s in total) before freezing onset.

2.6 *In vivo* electrophysiology

2.6.1 Single-unit recordings during fear conditioning and extinction

In vivo electrophysiological recordings during FC/Ext were performed at a collaborator's lab (Cyril Herry, INSERM Neurocentre Magendie, Bordeaux, France) by the same experimenter. The auditory fear conditioning and extinction paradigm was similar to the one described above, with slight modifications regarding the behavior setup as well as the tone presentations. Here, a commercial fear conditioning setup was used (Imetronic, Pessac, France), consisting of a behavior box, an electric grid floor, a light and a sound source located in a sound proof chamber. Two different chambers were used to create distinct contexts depending on the stage of experiment (context A and B). In context A, the behavior box was rectangular, the floor was either covered with black plastic (during habituation and recall) or a grid floor (during FC) and the box was cleaned before and after each session with 80% ethanol. In context B, the behavior box was a cylinder, the floor was covered with transparent plastic, and the cleaning agent was 1% acetic acid. To score freezing behavior, an automated infrared beam detection system located on the bottom of the experimental chambers was used (Coulbourn Instruments). The animals were considered to be freezing if no movement was detected for at least 2 s. Further the CS+ used in these experiments were pips of 7 kHz, 250 ms, repeated at 2 Hz for 30 s and the CS- were white noise pips, 500 ms, repeated at 1 Hz for 30s. Lastly, during extinction training, four CS- were included that were intermingled with the CS+ in the first part of the experiment, leading to a total experiment duration of 30 min.

To record single-unit activity in the pIC during each stage of the experiment, the implanted electrodes were connected to a headstage (Plexon) containing sixteen unity-gain operational amplifiers. Each headstage was connected via a motorized rotary joint (Imetronics, France) to a 16-channel PBX preamplifier (gain 2000x, Plexon) with bandpass filters at 300 Hz and 8 kHz. Spiking activity was digitized at 40 kHz and isolated by time-amplitude window discrimination and template matching using an Omniplex system (Plexon).

At the conclusion of the experiment, electrolytic lesions were administered before transcardial perfusion to verify electrode tip location.

2.6.2 Single-unit optrode recordings

For functional verification of optogenetic inhibition, eNpHR3.0-expressing mice implanted with and optrode were tethered to an optical patch cord, connected to a yellow laser (593 nm) and to an electrophysiological head stage (RHD 16ch, Intan Technologies). Animals were placed in a behavioral arena with transparent plexiglass walls and behavior was recorded with a webcam at 30 fps. Neuronal activity data were band-pass filtered between 0.5 and 7.5 kHz and digitized at 30 kHz with a 16-channel amplifier system (RHD2000, Intan Technologies). Mice received five pulses of constant yellow (593 nm) light for 40 s with an ISI duration of 20 s.

2.6.3 Single-unit data analyses

Spike sorting was performed with Offline Sorter 4.4.0 software (Plexon) and data were analyzed with Neuroexplorer 5.112 (Nex Technologies) and custom written MATLAB programs. Waveforms were manually defined with guidance from three-dimensional visualizations of the principal components, timing and voltage features of the waveforms. A single unit was defined as a cluster of waveforms constituting a discrete, isolated cluster in the feature space that did not contain spikes with a refractory period of less than 1 ms based on autocorrelation analyses. Unit isolation was verified with auto- and cross-correlation histograms, and data were excluded when either one unit was recorded on multiple channels, or when two or more units were recorded on the same channel.

Single unit recordings during FC/Ext

Event related changes in firing rates: Spike train activity from the isolated units was analyzed by first concatenating all recorded units from all mice within each session into a 'population vector', resulting in 93 units recorded in habituation, 82 in recall, and 93 in extinction. Note that for these experiments, I could not acquire data during FC, as the electrifiable grid floor used for US delivery interfered with electrophysiological recordings. To analyze CS+ and CS- evoked changes in firing rate, peri-stimulus time histograms (PSTHs) were computed for each neuron with 3 s bins and a z-score was calculated by comparing the firing rate after CS onset with the firing rate during the 30 s before CS onset. Neurons were defined as significantly inhibited or activated by the CS if their z-score was below or above 1.96 SD, respectively, in three or more bins following CS onset. To analyze freezing (i.e. 2 s after freezing event onset, 0 to 2 s) or pre-freezing (i.e. 2 s after pre-freezing event onset, -2 to - 0 s) evoked changes in firing rate, I analyzed data from the recall session (session with highest average amount of freezing) and also concatenated all recorded units from all mice into a 'population vector', resulting in 82 neurons, from

which one unit had to be excluded due to a z-score that exceeded 2xSD. I also excluded freezing events where the inter-freezing-interval was less than 6 s. Next, peri-stimulus time histograms (PSTHs) were computed for each neuron with 0.1 s bins and a z-score was calculated by comparing the firing rate after CS onset with the firing rate during the -6 to -4 s before freezing onset. Neurons were defined as significantly inhibited or activated by freezing if their z-score was below or above 1.96 SD, respectively, in five or more bins following freezing onset. Further, neurons were defined as significantly inhibited or activated by pre-freezing if their z-score was below or above 1.96 SD, respectively, in five or more bins following pre-freezing onset.

Fear state neurons: To identify fear state modulated neurons, I performed population-based analyses. This approach was suitable for my paradigm as mice were tested on three consecutive days, each session with a broad timescale and controlled stimuli. Additionally, although the variability of freezing behavior within each session was minimal, freezing between sessions was heterogeneous among mice (**Figure 13**) and therefore averaging neuronal activity across subjects at fixed time points may not produce easily interpretable results. To this end, I first extracted units that could reliable be identified across all stages of the FC/Ext paradigm (habituation, recall and extinction; note that none of those neurons could be reliably followed until extinction retrieval). To identify the same unit across different sessions, I cross-correlated the waveforms of units that were recorded on the same channel on different recording days. Only when the correlation coefficient r was ≥ 0.99, a unit was classified as the same across sessions. Spike train activity from those pursuable units was concatenated across the sessions (habituation, recall, and extinction) for population analyses. For each unit, the instantaneous spike count was temporally smoothed with a sliding-window of 5 s (0.1 s steps) and binned into 5 s bins. The population vector was then normalized using a z-score, where as baseline, the first 180 seconds of habituation were used. 'Fear state neurons' were classified as either 'fear state activated' or 'fear state inhibited' if their z-score was above or below 1.96 SD, respectively, in five or more bins during the first 180 s of recall (when average fear behavior was highest). For correlations between average activity of fear state neurons with freezing behavior, I used average z-scores and freezing probabilities of 15 consecutive time bins (each 250 s long) along the whole population vector.

Optrode recordings

To validate optogenetic pIC inhibition, peri-stimulus time histograms (PSTHs) were computed for each neuron with 1 s bins. A z-score was then calculated by comparing the firing rate during light application (40 s) with the firing rate during the 20 s before light onset and the z-score was averaged across the five

light pulses. Neurons were defined as significantly inhibited or activated by light if their z-score was below or above 1.96 SD, respectively, in five or more bins following light onset.

2.7 Vagus nerve stimulation (VNS) experiments

2.7.1 VNS during extinction learning

At the beginning of each experiment, the stimulation connector on the animal's head was tethered to an isolated electrical stimulator (Model 4100, A-M Systems) that was synchronized via TTL pulses to the ANYmaze software and the animal was placed in the fear conditioning box. The behavioral protocol for FC/Ext was similar to the standard protocol I used in my previous experiments, but had to be adapted since previous studies have reported that VNS effects on extinction were progressively developing across several days of VNS exposure (Noble et al., 2017). Thus, animals were undergoing three consecutive days of extinction, where the first two CS+ presentations were not paired with VNS, followed by ten CS+ presentations that were paired with VNS (31s duration (starting 1 s before CS+ onset), 0.4 mA, 30 Hz, 500µs pulse width). Accordingly, the first two unpaired CS+ of extinction day 1 served as fear memory recall (and thus also as the fear state of the animal), and the unpaired CS+ of extinction day 2 and 3 served as tests for extinction retrieval. As controls ('Sham'), I used animals that underwent the same VNS surgery, but never received any stimulation of the vagus nerve.

2.7.2 Effects of VNS on pIC activity

To test for general effects of VNS on pIC activity, I monitored pIC activity using fiber photometry while an animal was receiving a series of VNS ('VNS only'). The animal was placed in a behavioral arena and the stimulation connector on its head was tethered to an isolated electrical stimulator (Model 4100, A-M Systems) that was synchronized via TTL pulses to the ANYmaze software. Additionally, the fiber-optic implant was connected to the fiber photometry systems as described above. After an acclimatization period of 5 min, ten VNS bouts (0.4 mA, 500 µs pulse width and 30 Hz, for 31 s) were applied with a random ISI between 30 and 90 s. pIC activity was recorded throughout the whole experiment.

In the same animals, I recorded pIC population activity while they were undergoing the FC/Ext experiment, where VNS was paired with CS+ presentations during extinction (using the same stimulation parameters as described above), and pIC activity was recorded throughout the whole experiment. The

paradigm used here was similar as described above, with the exception that I stopped the experiment after the third recall session because both groups were already showing a complete extinction of CS+ evoked fear behavior (**Figure 25**). Thus, only two instead of three days of recall/extinction were conducted.

2.8 Histology and validation of injection and implantation sites

Animals were anesthetized with a mixture of Ketamine and Xylazine (100 mg/kg and 20 mg/kg bodyweight, respectively, Serumwerk Bernburg) and perfused intra-cardially with 4% paraformaldehyde (PFA) in phosphate buffered saline (PBS). Brains were post-fixed for an additional 24 h in 4% PFA at 4 °C. Then, coronal sections were cut at 70 µm with a VT1000S vibratome (Leica Biosystems). I collected all sections between +0.14 mm and -1.34 mm distance from Bregma to cover the whole pIC as well as the spread of virus. Sections were mounted onto glass slides using a custom-made mounting medium containing Mowiol 4-88 (Roth, Germany) as described elsewhere with 0.2 mg/mL DAPI (Sigma-Aldrich, MO). Images of the sections were acquired with an epifluorescent microscope (Leica DFC7000 T). The placement of the optic fibers as well as the spread of the viral infection was determined using the mouse brain atlas from Franklin & Paxinos, 4th edition[217]. For optogenetics and fiber photometry, only animals with a correct fiber placement and a viral spread confined to pIC were included in the study. For *in vivo* electrophysiology experiments, only animals with a correct electrode placement in the pIC were included in the study.

2.9 Statistical analysis

Analyses were performed with GraphPad Prism (GraphPad Software, version 8) and MATLAB (Mathworks). Group comparisons were done by either one-way or two-way ANOVA followed by Bonferroni corrected *post hoc* tests if a significant main effect or interaction was observed ($p < 0.05$).. Single-variable comparisons were made with two-tailed unpaired or paired t-tests ($p < 0.05$). Significance of averaged z-scores was tested using one-sample t tests with zero as theoretical mean ($p < 0.05$). Correlation analyses were made using linear regressions and differences between slopes were compared using ANCOVA. Asterisks or hashtags in the figures represent the p-values of *post hoc* or t tests corresponding to the following values *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ based on mean ± SEM, or accordingly #$p < 0.05$ and ##$p < 0.01$ based upon mean ± SEM. Numbers of mice for all experiments are reported in the figures and their legends. Data distribution was assumed to be normal, but this was not

formally tested. No statistical methods were used to predetermine sample size for single experiments, but the sample sizes were similar to or greater than those reported in previous studies related to our experiments. I excluded mice *post hoc* from optogenetic or photometry experiments when optic fiber placement or viral expression patterns were not appropriate (outside the target region, or when viral expression was poor) or when implants were lost during tests lasting for multiple days. All experiments reported here were reliably reproduced in independent experimental groups for behavior experiments and physiological recordings. For optogenetic experiments, littermates were randomly assigned to the different groups. Data collection and analysis were not performed with blinding to the conditions of the experiment. Computer-based analyses ensured unbiased data collection and analysis in most cases. Lastly, *post hoc* video scoring by human raters was performed without blinding.

3 RESULTS

3.1 Establishment of a fear conditioning and extinction paradigm

As a first step in my PhD project, I established a reliable and robust differential auditory fear conditioning and extinction (FC/Ext) paradigm that could be used for physiological and functional interrogations of the pIC.

To this end, mice were trained in a behavioral paradigm that consisted of five stages on four consecutive days. The paradigm started with a habituation on day 1, where the mice were exposed to two different auditory conditioned stimuli (CS) in context A that were 30 s long and played four times each. Two hours later, the mice were experiencing fear conditioning (FC) in the same context, where one of the CS (CS+) was five times coupled to a foot shock as an unconditioned stimulus (US), whereas the other CS was also played five times but never coupled to an aversive outcome (CS-). On day 2, fear memory was tested in a recall session by placing the mice again into context A and exposing them four times to both CS+ and CS- without the occurrence of a US. Extinction happened on day 3 in context B, where the CS+ was played twelve times without any aversive outcome, and extinction retrieval was tested on day 4 by playing the CS+ twice in extinction context B (**Figure 6b**). Conditioned fear behavior was measured as % time spent freezing (defined as complete absence of movement except breathing) during the CS+ or CS- (**Figure 6c).** During habituation, mice showed only very low levels of freezing behavior, although interestingly, freezing to the CS- seemed to be slightly higher in comparison to the CS+. Animals showed a robust increase in % freezing to the CS+ after FC that was also present during recall. In contrast, freezing evoked by the CS- only increased slightly during the initial phase of FC, but stayed below CS+ freezing during later phases of FC as well as during recall. When undergoing extinction, freezing evoked by the

CS+ slowly decreased again and was similar to starting levels at the end of extinction retrieval, indicating that the animals successfully extinguished the conditioned fear responses. Fear learning performance was quantified during recall (**Figure 6d)** by first comparing the % time spent freezing during CS presentations with % time spent freezing during 'preCS' (30 s preceding each CS) to control for fear generalization to the context. Indeed, animals reliably learned to associate the occurrence of the CS+ with an aversive outcome, indicated by a significant increase in freezing between preCS+ and CS+ (**Figure 6d left).** In contrast, there is not difference in freezing behavior between preCS- and CS-, indicating that the CS- does not evoke fear behavior (**Figure 6d middle).** Further, animals were also able to reliably discriminate between the CS+ and CS- as the % freezing to CS+ is significantly higher than to the CS- (**Figure 6d right).** Thus, a first conclusion from this experiment is that the differential auditory FC paradigm was leading to robust, reliable and specific fear learning. Next, I analyzed within session reduction in fear behavior during extinction by comparing the average CS+ evoked freezing during recall and during the last four CS+ presentations of extinction, as well as the long-term retention of extinction learning by comparing average CS+ evoked freezing during recall and during extinction retrieval (**Figure 6e).** Animals showed a significant reduction in freezing between recall and late extinction, as well as a significant difference in freezing between recall and extinction retrieval. Thus, animals successfully extinguished CS+ evoked conditioned fear responses at the end of the FC/Ext paradigm.

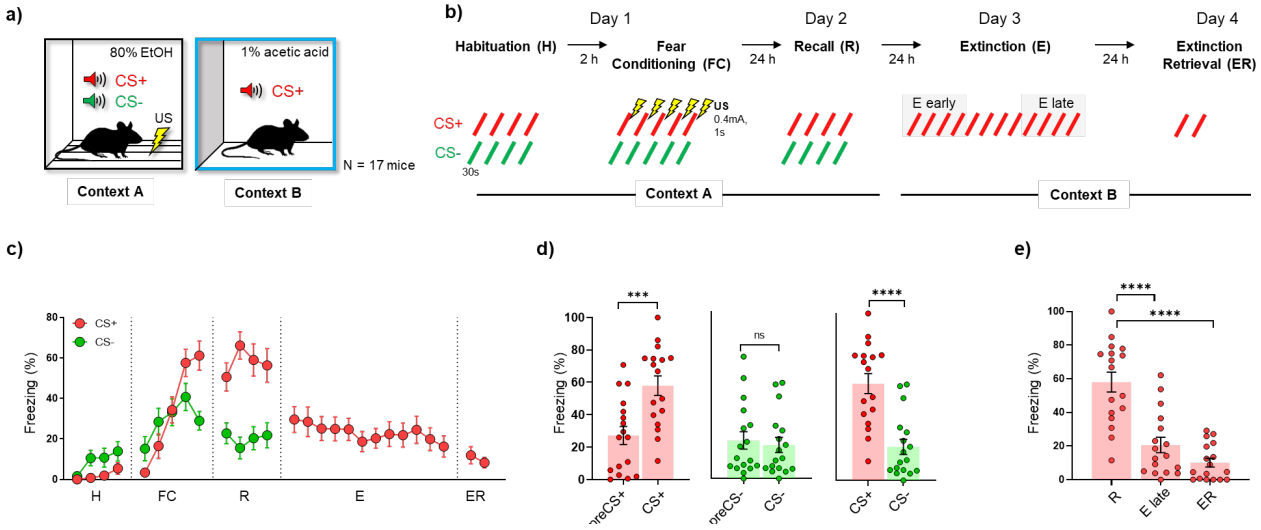


Figure 6: Quantification of fear behavior during fear conditioning and extinction.

a) *The experimental setup consisted of a behavior box where visual, olfactory and haptic cues were changed to create two distinct contexts. In context A, the box was equipped with a loud speaker and an electrifiable grid floor, and 80% ethanol was used as cleaning agent. In context B, the cleaning agent was 1% acetic acid, the grid floor was covered with plastic and an additional visual pattern was added to the walls. A total of N=17 mice was tested.* ***b)*** *The auditory fear conditioning and extinction protocol started with a habituation (H) on day 1, where the mice were exposed to two different auditory conditioned stimuli*

(CS) in context A that were 30 s long and played four times each. Two hours later, the mice were undergoing fear conditioning (FC) in the same context, where one of the CS (the CS+) was five times co-terminated with a foot shock that served as an unconditioned stimulus (US), whereas the other CS was also played five times but never coupled with an aversive outcome (CS-). On day 2, fear learning performance was tested in a recall (R) session by placing the mice into context A and exposing them four times to both CS+ and CS- without the occurrence of a US. Extinction (E) happened on day 3 in context B, where the CS+ was played twelve times without any aversive outcome, and extinction retrieval (ER) was tested on day 4 by playing the CS+ twice in extinction context B. ***c)*** *Conditioned fear responses to the CS+ were measured using freezing and expressed as % time spent freezing during CS presentations. Animals showed a robust increase in %freezing to the CS+ after FC that was also present during recall. In contrast, %freezing to the CS- increased slightly during the initial phase of FC, but stayed below CS+ freezing during later phases of FC and recall. During extinction, freezing to the CS+ slowly decreased again and returned to starting levels at the end of extinction retrieval.* ***d)*** *Fear learning performance was assessed during recall. Animals show a significant increase in % time freezing averaged across all four CS+ in comparison to freezing during preCS+ (preceding 30 s) (two-tailed paired t test, t=4.498, df=16, p***=0.0004), whereas no significant difference was found between averaged preCS- and CS- freezing (two-tailed paired t test, t=0.6458, df=16, p=0.5276). Freezing to CS+ was significantly higher in comparison to freezing during CS- (two-tailed paired t test, t=5.528, df=16, p****<0.0001).* ***e)*** *Comparison between average CS+ evoked freezing during recall and late extinction (last four CS+ presentations, E late) or extinction retrieval, respectively. Freezing significantly decreased upon extinction learning and was also still significantly lower during extinction retrieval (one-way RM ANOVA, F(1.624, 25.99)=39.79, ****p<0.0001; Bonferroni corrected post-hoc analyses revealed significant differences between recall and extinction late, ***p=0.0001 , and between recall and extinction retrieval, ****p>0.0001). Data are plotted as mean ± SEM.*

Taken together, I successfully established a differential auditory FC/Ext paradigm that reliably evoked and extinguished conditioned fear behavior within just four days, which made it a well-suited tool to study the involvement of the pIC in acquisition and extinction of learned fear behaviors in mice.

3.2 Fiber Photometry Recordings in pIC during FC/Ext

In a first step to characterize the neuronal activity in the mouse pIC when animals were undergoing FC/Ext, I used fiber photometry (FP) to record bulk neuronal activity from the pIC in awake behaving animals. Therefore, I unilaterally infected the pIC with an AAV coding for a calcium indicator under the CaMKIIα promotor (AAV9.*Camk2a*.GCaMP6s.WPRE.SV40) and implanted an optic fiber above the injection site (**Figure 7b**). I measured bulk real-time calcium transients as a proxy for neuronal activity of excitatory pIC neurons in mice across the entire FC/Ext paradigm (**Figure 7a and c)**.

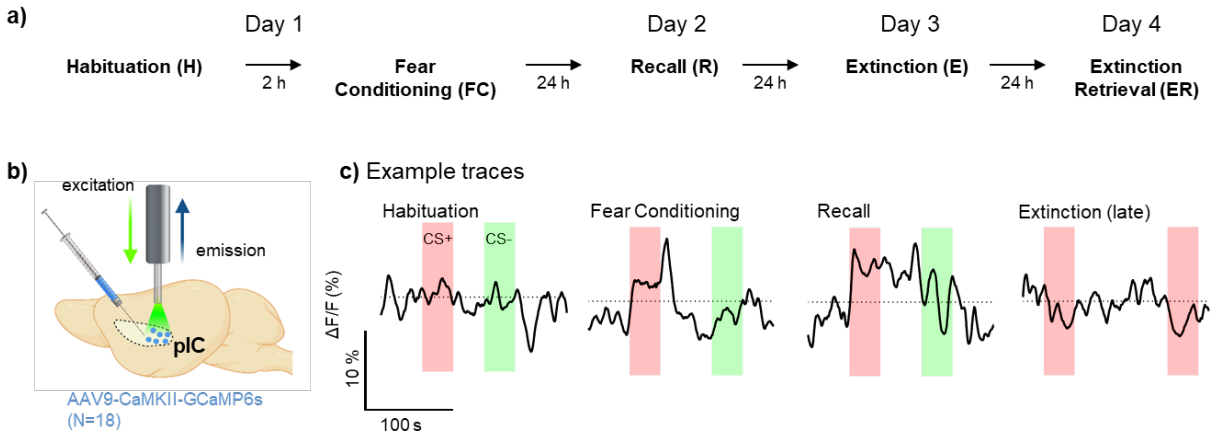


Figure 7: Fiber Photometry (FP) recordings in pIC during fear conditioning and extinction.
a) *The same fear conditioning and extinction protocol was used as in Figure 6b.* ***b)*** *Principle of fiber photometric recordings, optic fiber placement and virus used. A total of 18 mice was tested in FP experiments.* ***c)*** *Example GCaMP6s signals recorded from pIC excitatory neurons during FC/Ext. CS+ and CS- presentations are shaded in red and green, respectively. Note that CS+ exposure evokes an increase in pIC activity during fear conditioning and recall that is not present during habituation or late extinction.*

I analyzed cue evoked pIC activity across the entire FC/Ext paradigm. **Figure 8** displays average pIC excitatory neurons activity during 30 s of CS+ and CS- presentations across the different sessions of the experiment. Traces represent averaged z-scored activity to the first and last CS presentation within each session (**Figure 8a** and **c**). For CS+ evoked pIC responses, I found that during habituation, when the CS+ was not yet associated with a specific outcome, there was no significant change in pIC activity upon CS+ presentations (**Figure 8a** left). However, during the learning phase of the FC experiment when the CS+ was repeatedly coupled to the US, the CS+ significantly increased pIC activity, which was also still apparent during recall (**Figure 8a** and **Supplementary figure 2**; note that the US evoked activity will be discussed in **Figure 9)**. During extinction, this CS+ evoked activity rapidly decreased again and stayed low during extinction retrieval (see also **Supplementary figure 2)**. In contrast, the CS- evoked a small but significant pIC activity increase during the first CS- presentation during habituation (**Figure 8c** and **Supplementary figure 2**), but was not causing any significant responses in pIC during the remaining CS- presentations during habituation nor during FC. Interestingly, during recall, the second CS- was even causing a significant decrease in pIC activity (**Supplementary figure 2).**

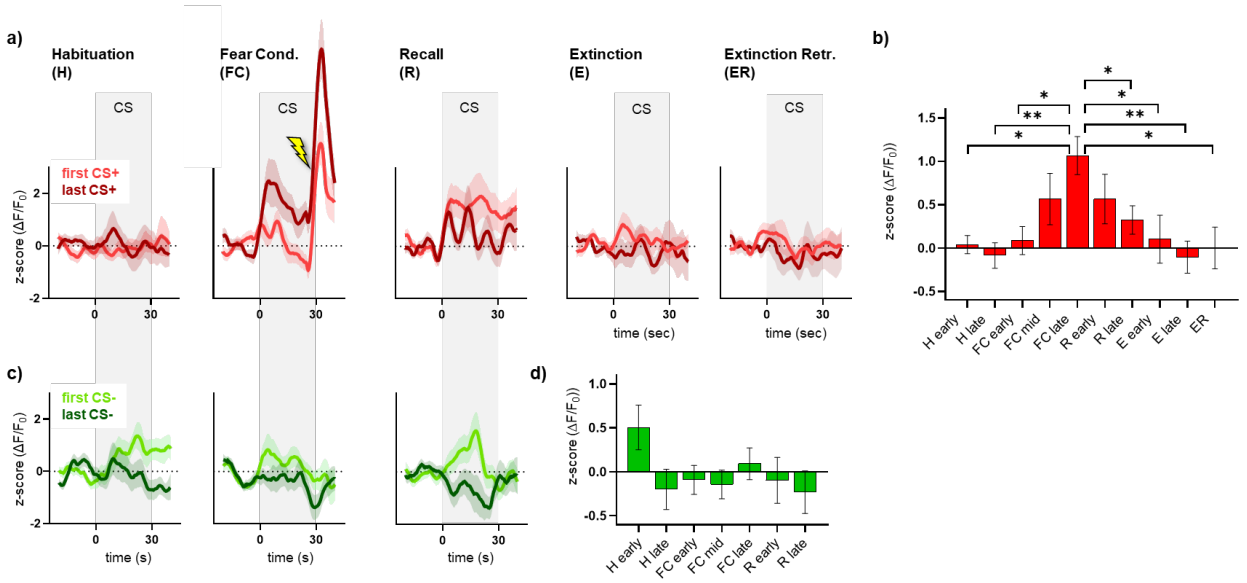


Figure 8: Cue evoked pIC activity is modulated by fear-eliciting cues across fear conditioning and extinction.
a) *pIC excitatory neurons activity during 30 s of CS+ presentations across the different sessions of the experiment habituation (H), fear conditioning (FC), recall (R), extinction (E), and extinction retrieval (ER). Traces are showing averaged z-scored activity to the first and last CS+ presentation in each session, grey shadings indicate 30 s of CS+ presentation (N = 18 mice).* ***b)*** *Bar graphs show average z-scores of two consecutive CS+ presentations (except for FC mid, where only one CS+ is shown). Average CS+ evoked pIC activity significantly and gradually increased during FC, and decreased again during recall and extinction (N=18 mice, one-way RM ANOVA, F(9,153)=3.454, p***=0.0007, Bonferroni corrected post hoc tests detected significant differences between habituation early and FC late, p*=0.01192, habituation late and FC late, p*=0.0037, FC early vs FC late, p*=0.0339, FC late and extinction early, p*=0.0410, FC late and extinction late, p**=0.0029, FC late and extinction retrieval, p*=0.118).* ***c)*** *pIC excitatory neuron activity during 30 s of CS- presentations across the different sessions of the experiment (habituation, FC, recall). Traces are showing averaged z-scored activity to the first and last CS+ presentation in each session, grey shadings indicate 30 s of CS- presentation (N = 18 mice).* ***d)*** *Bar graphs show average z-scores of two consecutive CS- presentations (except for FC mid, where only one CS- is shown). Average CS- evoked pIC activity is not significantly modulated during FC/Ext. (N=18 mice, one-way RM ANOVA, F(3.409,57.95)=1.613, p=0.1910). All data are plotted as mean ± SEM.*

Further, when I compared the average cue evoked pIC activity across the sessions, I found that the CS+ evoked activity was gradually increasing during FC and gradually decreasing again during extinction (**Figure 8b)**. This was indicated by a significantly higher pIC responses during the last two CS+ of FC (FC late) in comparison to pIC responses during the entire habituation the first two CS+ of FC (FC early), the last two CS+ of recall (R late), and the entire extinction and extinction retrieval sessions (**Figure 8b)**. In contrast to that, the average CS- evoked pIC activity did not significantly change across the phases of the FC/Ext paradigm (**Figure 8d)**.

Further, I could also observe a strong response to the US itself (**Figure 9b)**, which is also in line with several studies showing that the IC responds to noxious somatosensory stimuli and thus, the IC is known to play an important role in pain processing [218,219]. Interestingly, amplitudes of foot shock evoked responses significantly increased between FC early (first two US presentations) and FC late (last two US presentations) (**Figure 9b)**.

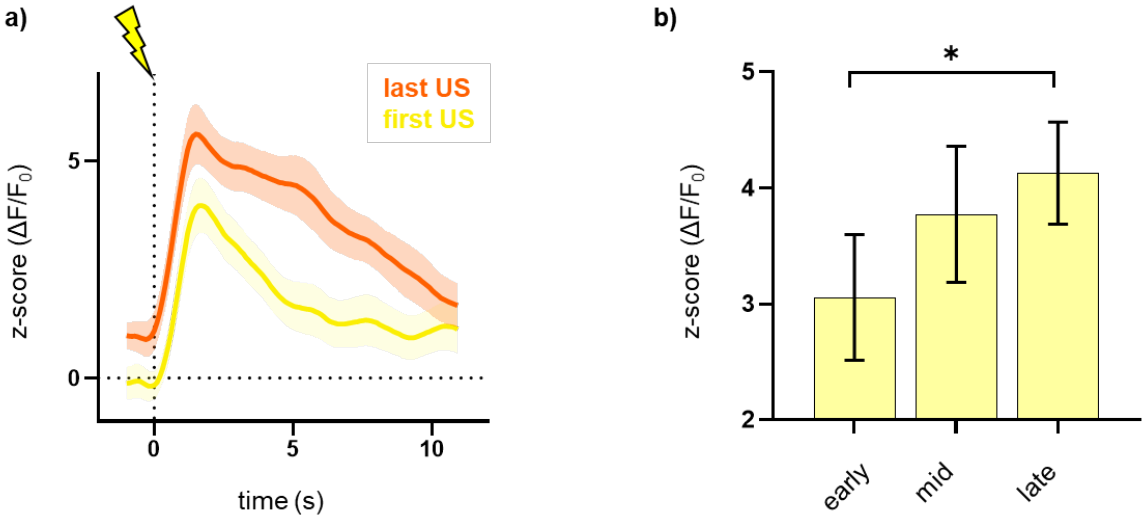


Figure 9: US evoked pIC activity gradually increases during FC.
a) *GCaMP6s traces are showing average z-scored pIC excitatory neurons activity to the first (yellow) and last (orange) US presentation during FC. The dotted vertical line indicates the onset of the US.* ***b)*** *Average US evoked pIC activity significantly increased between early FC (first two US presentations) and late FC (last two US presentations) (N=18 mice, one-way RM ANOVA, F(1.949, 33.14)=3.979, p*=0.0292; Bonferroni corrected post hoc tests reveal a significant difference between FC early and FC late, p*=0.0254). Data are plotted as mean ± SEM.*

Next, I wanted to investigate whether the observed gradual modulation of cue evoked pIC activity was also mirrored in the fear behavior of the animals. Therefore, I correlated average CS+ evoked neuronal activity with average CS+ evoked freezing across the entire FC/Ext experiment (**Figure 10**). Interestingly, freezing and pIC activity seemed to gradually increase and decrease in parallel (**Figure 10a)**. Linear regression revealed as strong positive correlation between pIC activity and fear behavior (**Figure 10b)**, leading to the conclusion that the pIC is highly engaged in states of elevated fear. In contrast, CS- evoked pIC responses and fear behavior did not show such a correlation: even though the CS- evoked a small but

considerable amount of freezing after FC, CS- related pIC activity remained stable throughout the sessions of the FC/Ext experiment (**Supplementary figure 3**).

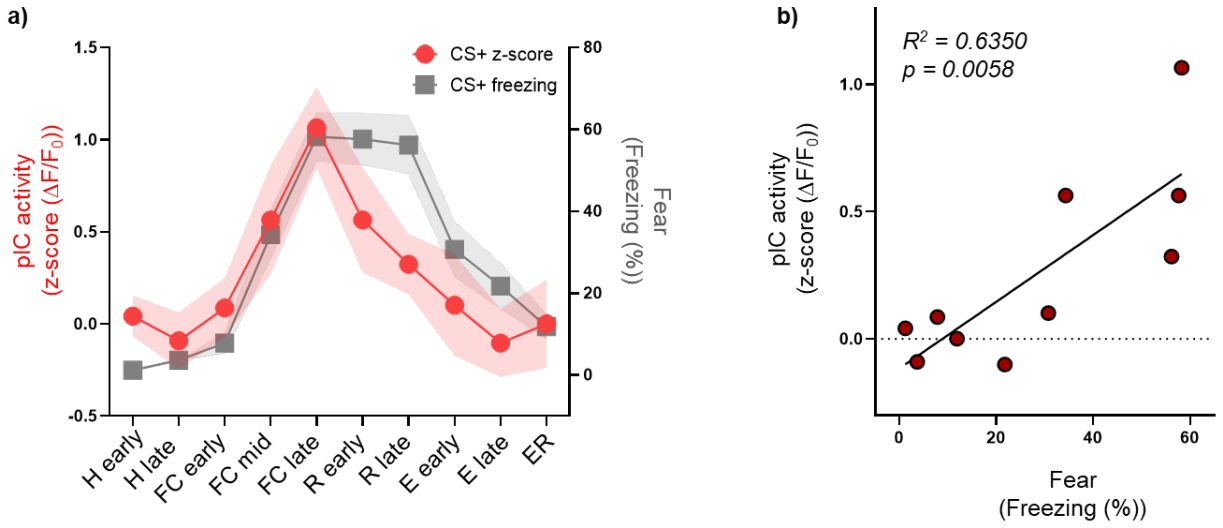


Figure 10: pIC activity correlates with fear levels.
a) Fear cue evoked pIC excitatory neurons activity and freezing both gradually in- and decreased across the sessions of the paradigm. Values are shown as average z-score or % freezing during two consecutive CS+ presentations, respectively (except for FC mid, where values for only one CS+ is shown). N=18 mice, data are plotted as mean ± SEM. **b)** Same data as in **a)** plotted using a linear regression, which revealed a strong positive correlation between CS+ evoked fear behavior and pIC activity (N=18 mice, $F(1,8)=13.92$, $R^2=0.6350$, slope significantly non-zero, p**=0.0058).

The observation that pIC responses were amplified with increasing fear levels raised the question whether the pIC was just indirectly representing freezing behavior. To exclude that possibility, I specifically analyzed freezing evoked pIC activity and surprisingly, I found that 2 s before a freezing event, pIC activity showed a slow but apparent increase, before it significantly and rapidly decreased during the freezing event itself (**Figure 11a)**. Whereas the rise in activity that preceded freezing events was only close to significant, the activity drop during freezing was highly significant (**Figure 11b)**. Thus, the observed increase in fear-cue evoked pIC responses during phases of high fear cannot be simply explained by an increase in freezing behavior.

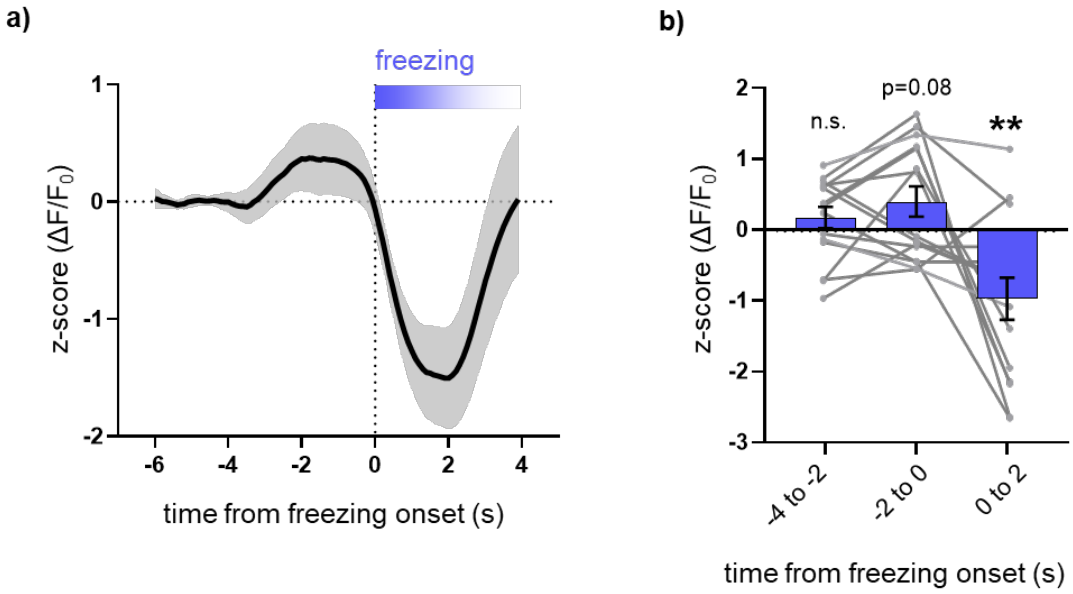


Figure 11: pIC population activity decreases during freezing.

a) *GCaMP6s trace shows average z-scored pIC activity of excitatory neurons aligned to all freezing events (that are at least 2 s long) during recall. Note that 2 s before freezing, pIC activity slowly increased before it significantly and rapidly decreased during the freezing event itself. Vertical dotted line indicates freezing onset. Data are normalized to -6 to -4 s before freezing onset (data were cleaned to make sure that no freezing occurs within this time window). N=15 mice, data are plotted as mean ± SEM.* ***b)*** *Average z-score of three different 2 s long time bins around freezing onset reveals that the increase 2 s before freezing was close to significant, whereas pIC activity significantly decreased during the freezing event (N=15 mice, two-sided one sample t tests with zero as theoretical mean, -4 to -2 s bin: t=1.150, df=14, p=0.2694, -2 to 0 s bin: t=1.879, df=14, p=0.0812, 0 to 2 s bin: t=3.276, df=14, p**=0.0055). Grey lines indicate the values for single animals. All data are plotted as mean ± SEM.*

In conclusion, using FP recordings of excitatory pIC neurons, I could show that pIC representations of qualitatively identical auditory stimuli changed after they were associated with a different valence upon fear learning, and went back to baseline levels upon extinction learning. Further, also painful stimuli were differentially processed when the animal entered a more fearful state. Collectively, this suggests that pIC activity represents individual fear levels as I could show that the pIC processes sensory stimuli, in a (fear-) state-dependent manner.

3.3 *In vivo* single unit recordings in pIC during FC/Ext

The findings described above provide evidence that populations of pIC neurons state-dependently process sensory stimuli. However, recording neuronal activity via calcium dynamics using fiber photometry is only a proxy for actual neuronal activity and can solely measures the net neuronal activity and cannot provide information about the activity pattern of different subpopulations. Additionally, due to the slow dynamics of genetically encoded calcium indicators, GCaMP6s has a temporal resolution that is too slow to capture actual neuronal firing that happens within millisecond range [220,221]. Therefore, I next used *in vivo* electrophysiology to record from single neurons in freely behaving mice with high spatial and temporal accuracy.

3.3.1 Setup for *in vivo* single unit recordings in pIC during FC/Ext

In order to acquire electrophysiological data from mice undergoing FC/Ext, I carried out single unit recordings in Bordeaux in the lab of a collaborator. I used self-manufactured 16-wire electrode arrays that were implanted into the mouse pIC (**Figure 12a)**. To simultaneously record neuronal activity and mouse behavior during FC/Ext, I used an *OmniPlex* System (Plexon, Texas, US**A)** synchronized with a behavioral setup from Imetronic (Pessac, France) (**Figure 12b, Figure 13a)**, respectively. I used well-established methods for isolating the spiking activity from the wide band signal as well as for sorting of single unit activity using *PlexControl* and *OfflineSorter* software, respectively (Plexon, Texas, USA) (**Figure 12b and c)**. As an example for pIC neuronal activity during FC/Ext, **Figure 12d** shows peri-event raster plots and histograms of two different representative neurons during four consecutive auditory cue presentations, where the neuron on the left shows a reduction in firing, whereas the neuron on the right increases its activity during the cue.

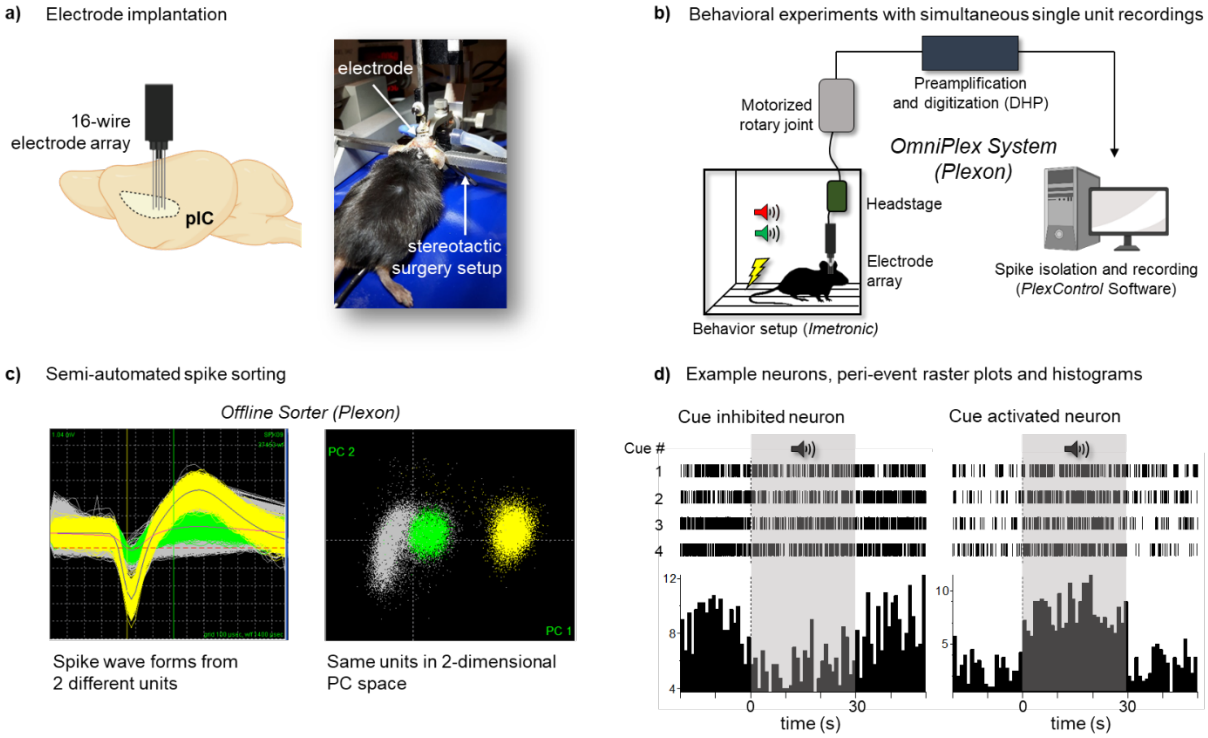


Figure 12: Experimental approach for in vivo single unit recordings in pIC during fear conditioning and extinction.
a) *A custom-made 16-wire electrode array was stereotactically implanted to unilateral pIC.* ***b)*** *Electrophysiological data were acquired using an OmniPlex system (Plexon). The experimental setup consisted of a behavior box with distinct context, a mouse with the implanted electrode array that was connected via a headstage to a motorized rotary joint to reduce torque, a Digital Headstage Processor (DHP) that in turn sent the preamplified and digitized electrophysiological data to a computer where an initial thresholding and spike data isolation step happened using the PlexControl software.* ***c)*** *Example for semi-automated spike sorting using Offline Sorter software (Plexon). Left: Extracellular waveform data from two isolated units in the same channel (green and yellow traces: isolated units, grey: multiunit background noise). Right: Principle component analysis of the same units as left. First two principle components are plotted revealing two well-isolated clusters, consisting of one separate unit each.* ***d)*** *Peri-event raster plots and histograms of two different example neurons during four consecutive auditory cue presentations (grey shaded area, bin size=1 s). The neuron on the left shows a reduction in firing during the cue, whereas the neuron on the right increases its firing rate during the cue.*

The behavioral paradigm used for FC/Ext was similar to the paradigm described above (**Figure 6b)**, with the exception that during extinction, I additionally presented four CS- in the beginning of the session (E early) (**Figure 13b)**. Analysis of fear behavior of mice used for electrophysiology experiments showed that the animals successfully performed differential fear and extinction learning (**Figure 13c**). Mice showed an increase in CS+ evoked freezing during FC that was highest during recall and decreased again upon extinction, whereas CS- evoked freezing remained mostly unaltered (**Figure 13c**). Analysis of fear learning performance indicate that fear learning was also specific to the CS+ as the mice showed a

significantly higher freezing to the CS+ in comparison to 30 s preceding the tone onset (preCS+) and to the CS- (**Figure 13d left and right**). Interestingly, mice were also freezing significantly less during CS- in comparison to preCS-, indicating that in this setup, the CS- might acted as a so called 'safety signal' that is able to reduce fear even below baseline levels [105,222,223]. Further, extinction caused a reduction in CS+ evoked fear behavior as indicated by significantly lower freezing levels during extinction late and extinction retrieval in comparison to recall (**Figure 13e**).

Taken together, the behavior results obtained in the single unit recording experiment were thus qualitatively comparable to those in FP experiments.

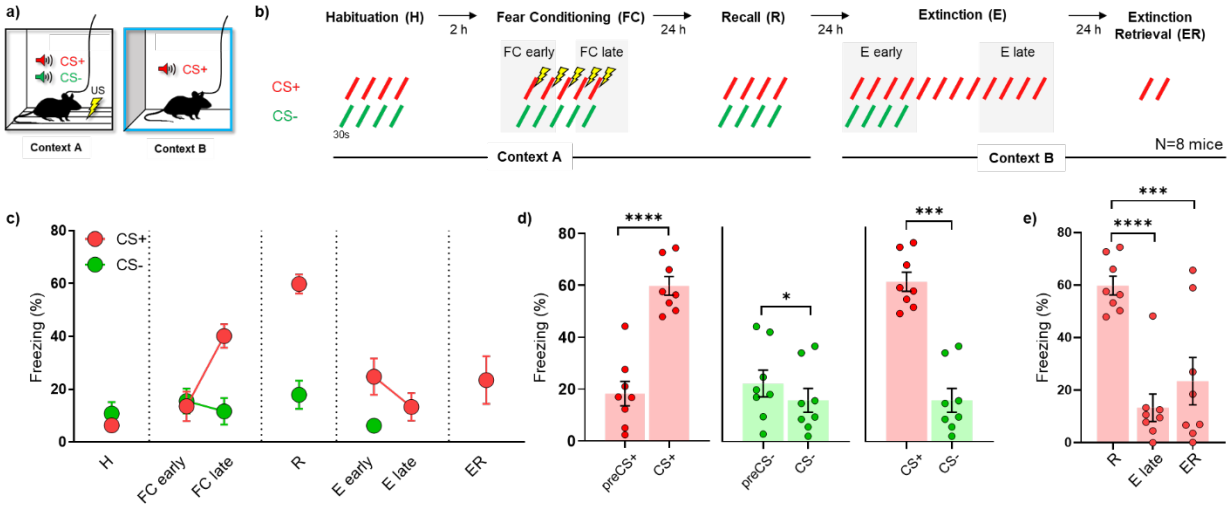


Figure 13: Experimental protocol and fear behavior of animals used for in vivo single unit recordings. ***a)*** *Behavioral setup for simultaneous single unit and behavioral recordings during FC/Ext.* ***b)*** *The FC/Ext protocol was similar as the protocol used before (Figure 6b), except that during early phases of extinction (E early), four CS- presentations were additionally interspersed during the beginning of the experiment.* ***c)*** *Fear behavior measured as % time spent freezing during CS+ and CS- during FC/Ext. Data are plotted as average of four CS presentations for habituation (H), recall (R), and extinction (E), and as average of two CS presentations for fear conditioning (FC) and extinction retrieval (ER) (N=8 mice). d) Analysis of fear learning performance during recall revealed that animals were freezing significantly more during CS+ in comparison to the 30 seconds preceding tone onset (preCS+) (two-tailed paired t test, t=9.631, df=7, p****<0.0001), whereas freezing was even significantly lower during CS- in comparison to preCS- (two-tailed paired t test, t=9.631, df=7, p****<0.0001). Animals successfully discriminated between CS+ and CS-, as CS+ freezing was significantly higher in comparison to CS- freezing (two-tailed paired t test, t=7.722, df=7, p***=0.0001). e) Extinction significantly decreased freezing behavior (one-way RM ANOVA, F(2, 14)=30.01, ****p<0.0001; Bonferroni corrected post-hoc analyses revealed significant differences between recall and extinction late, ****p<0.0001, and between recall and extinction retrieval, ***p=0.0001). Data are plotted as mean ± SEM.*

3.3.2 Cue and freezing evoked pIC single unit activity

Next, I analyzed cue and freezing evoked responses of all recorded pIC neurons within each session of the FC/Ext paradigm to test whether neurons showed a significant modulation in firing rate upon those events. In total, I recorded between 82 and 93 neurons from eight mice within each stage of the experiment (between 7 and 23 units per animal). Note that for these experiments, I could not acquire data during FC, as the electrifiable grid floor used for US delivery interfered with electrophysiological recordings. Further, in the following analyses I did not explicitly differentiate between neuron subtypes (i.e. pyramidal cells and interneurons), due to limited cell numbers (see Limitations and outlook).

In order to assess cue-evoked responses, I binned the firing rate into 3 s bins and normalized each neuron's firing rate to the 30 s preceding each CS. Units were classified as excited or inhibited by the CS if their z-score was above or below 1.96 SD, respectively, in three or more bins following tone onset (see **2.6.3**). I found that in each stage of the experiment, there were subpopulations of pIC neurons that were highly responsive to the CS+ (**Figure 14**), either showing a significant activation or inhibition upon CS+ presentations, which is in accordance to previous findings [97] (**Figure 14a and b)**. Interestingly, the percentage of CS+ responsive neurons out of all recorded units changed across the stages of the FC/Ext experiment: the fraction of both CS+ activated and CS+ inhibited neurons gradually increased upon FC and was highest during recall (the session where freezing/fear was highest, see **Figure 13c**), then gradually decreased again upon extinction learning, and was lowest during extinction retrieval, where also fear levels were low again. Remarkably, CS+ evoked pIC activity during extinction retrieval was even below initial values observed during habituation (**Figure 14c**). Further, the fraction of CS+ activated units was continuously higher than the CS+ inhibited fraction.

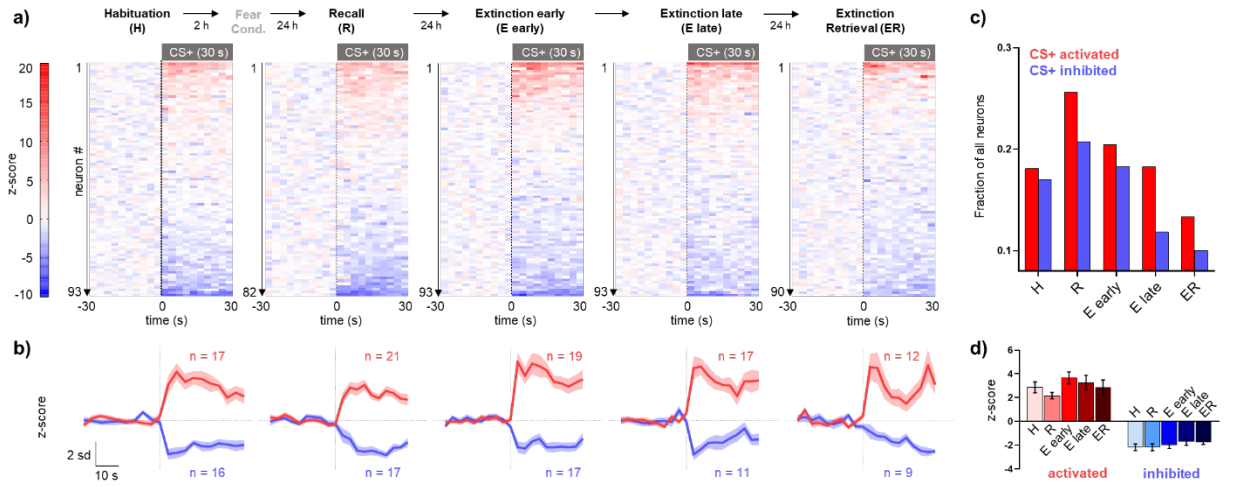


Figure 14: CS+ evoked pIC single unit activity.

a) *Top: Timeline of the experiment. (Note that data for FC were not collected due to electric noise artifacts during recordings on the electrifiable grid floor.) Bottom: Heatmaps of average z-scored neuronal responses from all units of all mice (N=8 mice, n=between 82 and 93 units per session). Responses to four CS+ presentations were averaged during habituation (H), recall (R), extinction early (E early) and extinction late (E late) and two CS+ presentations during extinction retrieval. Vertical dashed line indicates CS+ onset. Neurons were sorted by strongest responses to the CS+. Total number of recorded neurons in each stage is indicated on the left of each heatmap.* ***b)*** *Average population responses (z-score) of neurons that met the criterion for a significant change in firing rate upon CS+ presentation ('CS+ responsive neurons') across the different sessions of the experiment. The numbers of responsive neurons in each session are indicated on the graph. Data are plotted as mean ± SEM.* ***c)*** *Fractions of CS+ responsive neurons (activated or inhibited) out of all recorded neurons (pooled from all animals) for each session of the experiment. Note that the fraction of both CS+ activated and CS+ inhibited neurons was highest during high fear stages (recall) and lowest during extinction retrieval (even below initial values during habituation).* ***d)*** *Average response magnitude (z-score) of CS+ responsive neurons (activated or inhibited) during all sessions of the experiment. Note that there was no significant difference in response magnitude between the stages (one-way ANOVA, activated $F(4, 81) = 1.532$, $p = 0.2006$, inhibited $F(4, 65) = 0.5373$, $p = 0.7088$).*

Similarly, I found subpopulations in the pIC that were significantly responsive to the CS- cue during habituation, recall and extinction early (**Figure 15a and b**). However, the fraction of CS- responsive units across the three sessions did not follow the same pattern as the CS+ responsive units: while the CS- inhibited fraction slightly decreased between habituation, recall and extinction early, the CS- activated fraction seemed to first decrease between habituation and recall, but then increased again during extinction early (**Figure 15c**). Lastly, the amplitude of CS- responses remained unaltered across the stages of the FC/Ext experiment, similar to what I have observed in the CS+ responses (**Figure 15d**).

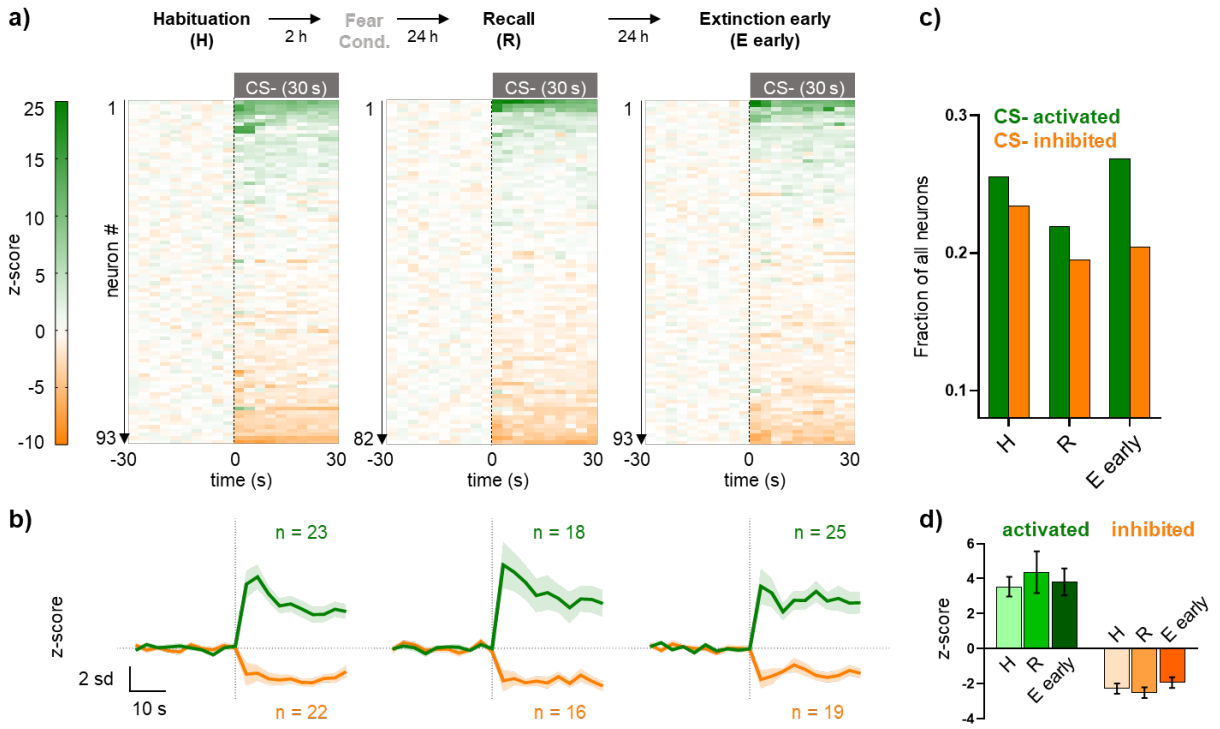


Figure 15: CS- evoked pIC single unit activity.
a) *Top: Timeline of the experiment. (Note that data for FC were not collected due to electric noise artifacts during recordings on the electrifiable grid floor. Bottom: Heatmaps of average z-scored neuronal responses from all units of all mice (N=8, n=between 82 and 93 units per session). Responses to four CS- presentations were averaged during all sessions. Vertical dashed line indicates CS- onset. Neurons are sorted by strongest responses to the CS-. Total number of recorded neurons in each sessions is indicated on the left of each heatmap.* ***b)*** *Average population responses (z-score) of neurons that met the criterion for a significant change in firing rate upon CS- presentation ('CS- responsive neurons') across the different sessions of the experiment. The numbers of responsive neurons in each sessions are indicated on the graph. Data are plotted as mean ± SEM.* ***c)*** *Fractions of CS- responsive neurons (activated or inhibited) out of all recorded neurons for each sessions of the experiment.* ***d)*** *Average response magnitude (z-score) of CS- responsive neurons (activated or inhibited) during all sessions of the experiment. Note that there was no significant difference in response magnitude between the sessions (one-way ANOVA, activated F (2, 63) = 0.6077, p = 0.7923, inhibited F (2, 54) = 0.8534, p = 0.4316).*

Further, I also compared the fractions of CS+ and CS- responsive units (out of all responsive cells) as well as units that were responsive to both cues across the sessions of the FC/Ext paradigm (**Figure 16**). During habituation, a larger fraction was responsive to the CS- (45%) in comparison to the CS+ (23%), whereas this was inversed during recall, where the CS+ responsive population (41%) was higher than the CS- responsive cells (33%). In extinction early, both fractions returned to values similar to habituation (CS+: 24%, CS-: 38%). Additionally, I found a substantial overlap between both CS+ and CS- responsive units, but the magnitude of the overlap was altered across the stages of the experiment and was the smallest

during recall (the session with the highest level of fear behavior). Conclusively, these observations indicate that the pIC seemed to augment the representation of fear-evoking cues during states of high fear.

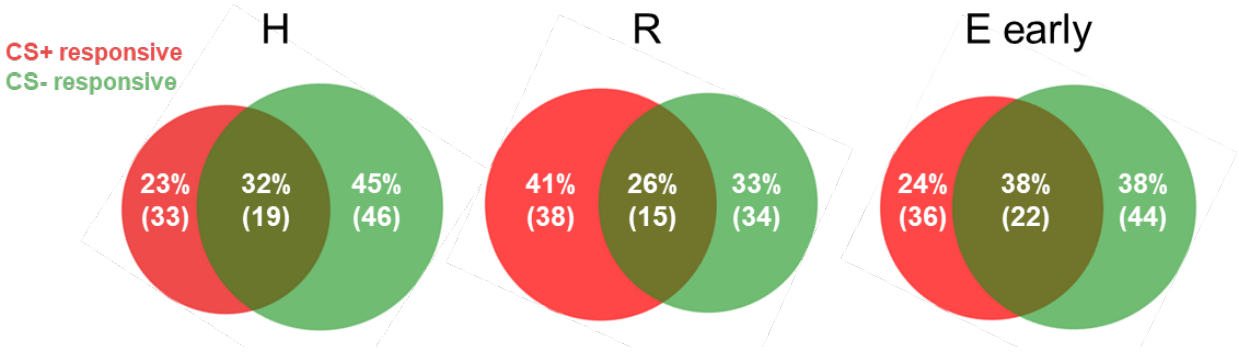


Figure 16: Fractions of CS+ and CS- responsive neurons across the different sessions and their overlaps.
Venn diagrams represent the fraction (out of all responsive units) for CS+ and CS- responsive neurons (activated and inhibited together) during habituation (H), recall (R) and extinction early (E early). During habituation, a larger fraction was responsive to the CS- (45%) in comparison to the CS+ (23%), whereas this was inversed during recall, where the CS+ responsive population (41%) was higher than the CS- responsive cells (33%). In extinction early, both fractions returned to values similar to habituation (CS+: 24%, CS-: 38%). Additionally, I found a substantial overlap between both CS+ and CS- responsive units, but the magnitude of the overlap was altered across the stages of the experiment and was the smallest during recall (the session with the highest level of fear behavior).

Taken together, these results confirm and extent the observations I made in fiber photometry-based activity changes throughout FC/Ext, where I saw a gradual increase in CS+ evoked activity with increasing levels of fear, a gradual decrease when fear is lessening again, as well as a mostly unaltered CS- response across the FC/Ext experiment. *In vivo* single unit recordings revealed that neurons did not change their CS+ response amplitude, but the amount of recruited pIC neurons increased between low and high fear states, with the CS+ activated fraction being larger than the CS+ inhibited fraction, thus resulting in a net increase of bulk pIC activity, similar to what was observed in fiber photometry recordings. Accordingly, CS- evoked activity did not follow a clear low - high fear state pattern, neither in fiber photometry, nor in electrophysiological recordings.

Further, having observed in fiber photometry recordings that excitatory neuronal pIC activity is not only modulated by fear-eliciting cues, but also by fear behavior (freezing), I here also analyzed neuronal firing rates around freezing events. I was especially interested to investigate the slow rise in pIC activity that happened roughly two seconds before freezing onset on a single neuron level, as a few studies have

already revealed that there are signatures in brain activity that can predict freezing onset [224,225]. To this end, I binned the firing rates into 100 ms bins, all freezing events from the recall session (where most freezing events happened) and normalized the neuronal firing rate to -6 to -4 seconds preceding freezing onset. I identified neurons that were responsive to either freezing itself (during the first two seconds of the freezing event) or during the preceding two seconds. Units were classified as excited or inhibited if their z-score was above or below 1.96 SD, respectively, in five or more bins following event onset (see **2.6.3**).

Indeed, I found a population of neurons that was significantly responsive during the freezing event itself ('freezing neurons', **Figure 17a and b)**, and another population of neurons that was showing a significant change in firing rate during the two seconds preceding freezing onset ('pre-freezing neurons', **Figure 17c and d)**. Surprisingly, in both populations, I found neurons that were significantly inhibited (supporting the findings from fiber photometry recordings) but also an almost equal number of neurons that was significantly activated **(Figure 17b and d)**. Freezing neurons (both the activated and inhibited group) showed a significant change in firing rate between the two seconds preceding freezing onset ('pre') and the two seconds during freezing ('during') (**Figure 17f left)**, whereas pre-freezing responsive neurons (again both the activated and inhibited group) did not change their firing upon freezing onset (**Figure 17f right)**. Further, when I analyzed the proportion of freezing and pre-freezing responsive units out of all recorded neurons, I found that around a quarter of all neurons belonged to the 'freezing' population, and around 15% belonged to the 'pre-freezing' group (**Figure 17e)**. Although there was a considerable overlap between those two populations (31% of responsive units are both freezing as well as pre-freezing responsive), still more than half (58%) were 'freezing' exclusive neurons, and 15% were 'pre-freezing' exclusive neurons (**Figure 17g)**.

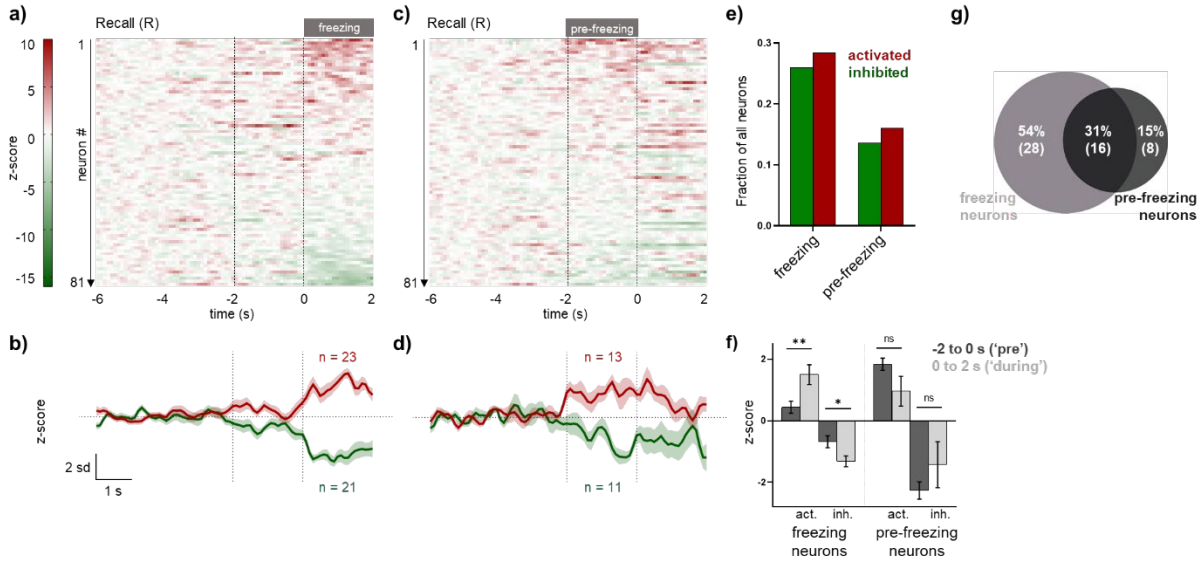


Figure 17: Freezing correlated single unit activity.
a) *and* ***c)*** *Heatmaps of average pIC neuronal responses around freezing onset from all units of all mice during recall (N=8 mice, n=81 units). Responses were averaged across all freezing episodes for each mouse and normalized to -6 to -4 s before freezing onset. The two vertical dashed lines indicate two seconds prior to freezing onset ('pre-freezing', -2 s) and freezing onset ('freezing', 0 s), respectively.* ***a)*** *Neurons were sorted by strongest freezing evoked response.* ***b)*** *Average population responses (z-score) of freezing responsive (activated or inhibited) pIC neurons. Data are plotted as mean ± SEM.* ***c)*** *Neurons were sorted by strongest pre-freezing response.* ***d)*** *Average population responses (z-score) of pre-freezing (activated or inhibited) pIC neurons. Data are plotted as mean ± SEM.* ***e)*** *Fractions of freezing and pre-freezing neurons (activated or inhibited together) out of all recorded neurons during recall.* ***f)*** *Average response magnitude (z-score) of freezing and pre-freezing neurons during 'pre' (-2 to 0 s) and 'during' (0 to 2 s) freezing periods. Left: Freezing neurons showed a significant change in firing rate upon freezing onset (two-tailed paired t tests for pre vs during; activated freezing neurons: t=5.939, df=22, p****<0.0001, inhibited freezing neurons: t=4.723, df=20, p***=0.0001). Right: In contrast, pre-freezing neurons did not change their firing upon freezing onset (two-tailed paired t tests for pre vs during; activated pre-freezing neurons: t=0.9226, df=12, p=0.3744, inhibited pre-freezing neurons: t=0.1625, df=10, p=0.8742). Data are plotted as mean ± SEM.* ***g)*** *Venn diagram representing the number of all freezing (light grey) and pre-freezing (dark grey) neurons during recall. The percent overlap is calculated from the total number of responsive neurons. Note that even though there is a substantial overlap, there are still sub-populations of freezing or pre-freezing specific neurons.*

Taken together, I identified a subpopulation of pIC neurons that encodes acute expression of fear behavior (freezing neurons), and another subpopulation that changes its activity shortly before the expression of fear behavior (pre-freezing neurons). Those results further corroborate the conclusion that the pIC is strongly engaged when fear levels are heightened.

3.3.3 Encoding of sustained fear states in pIC neurons

Using two-photon microscopy in awake head-fixed mice, our lab has recently found that a large fraction of pIC neurons exhibited greater activity during sustained phases of anxiety [35]. In order to better and more comprehensively understand the dynamics and nature of population coding of sustained anxiety states, I re-analyzed the electrophysiological data in order to potentially identify similar 'anxiety state' signatures in pIC single units. I hypothesized that there would be neurons showing a change in activity across a large time scale, following the rather slow transitions between low and high fear states exhibited by an animal across the FC/Ext paradigm. To this end, I first identified neurons that could reliably be followed across all stages of the FC/Ext paradigm (25 neurons from seven animals) and created a population vector by concatenating their normalized firing rate across three sessions of the experiment (habituation, recall and extinction; note that none of those neurons could be reliably followed until extinction retrieval) (see **2.6.3**) (**Figure 18**). I then identified 'fear state neurons', that were classified as either 'fear state activated' or 'fear state inhibited' if their z-score was above or below 1.96 SD, respectively, in five or more bins during the first 180 s of recall (when average fear behavior was highest) (**Figure 18a)**. Remarkably, those fear state neurons showed a long lasting change in activity when the mice were re-exposed to the fear-evoking context and cues during recall, and this activity change outlasted until the end of extinction (**Figure 18b**). Importantly, when I correlated the average activity of fear state neurons across time (in 250 s bins) with the respective freezing probability within the same time window, I found a strong correlation between pIC activity and fear behavior for both fear state activated as well as inhibited neurons (**Figure 18d**): the higher the freezing probability, the stronger the response amplitude of the state neurons. Further, when looking at the fraction of neurons that were identified as a fear state neuron, I found that a majority of the 25 recorded neurons were modulated by the fear state (32% were fear state activated, and 44% were fear state inhibited, **Figure 18c**), whereas only a small portion did not show a state-dependent modulation (24%), indicating that the pIC is highly engaged in processing sustained states of high fear.

These results strongly suggest that the pIC is not only state-dependently responding to learned fear cues, but also that a large fraction of pIC neurons exhibit a long-lasting modulation during states of high fear.

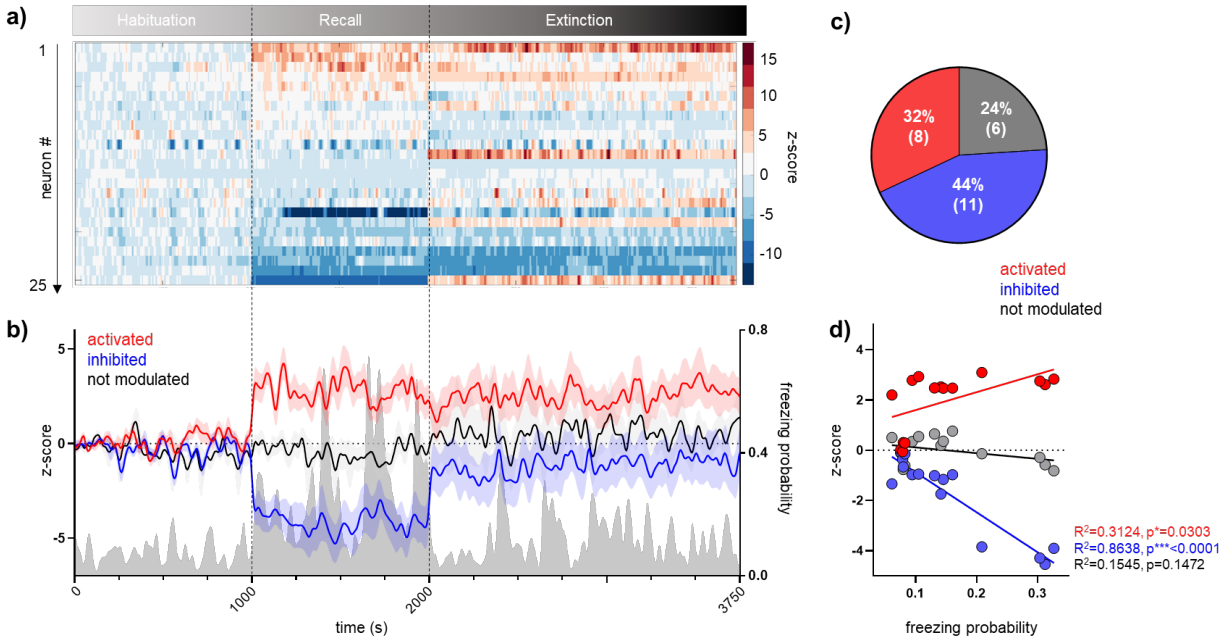


Figure 18: Long-lasting anxiety states are encoded in a subset of pIC neurons.
a) *Heatmap of pIC population activity concatenated for habituation, recall and extinction, where each row is one neuron, total duration is 3750 s. Only neurons were used that could reliably be identified across all three stages (n=25 neurons from N=7 mice). Activity was normalized to the first 180 s of Habituation and neurons were sorted according to response magnitude during the first 180 sec of recall. Vertical dashed lines indicate the start and end of each respective session.* ***b)*** *Average z-scored population activity of 'fear state neurons' (left y-axis), and corresponding average freezing probability shaded in grey (right y-axis). Fear state neurons were classified as 'activated' or 'inhibited' when their activity significantly increased (red) or decreased (blue) during the first 180 s of recall. Accordingly, 'not modulated' neurons did not show a significant change in activity. Fear state neurons showed a long-lasting change in activity during recall and extinction. Data are plotted as mean ± SEM.* ***c)*** *Fractions of activated and inhibited fear state neurons or not modulated neurons indicate that a majority (76%) of neurons were modulated during high fear states.* ***d)*** *Correlation between neuronal activity and freezing probability, each data point corresponds to a 250 s time bin (15 consecutive time bins in total). Linear regression revealed a strong correlation between fear state neuron activity and fear level (activated neurons: F(1,13)=5.907, R2=0.3124, slope significantly different from zero, p*=0.0303; inhibited: F(1,13)=82.44, R2=0.8638, slope significantly different from zero, p***<0.0001; not modulated: F(1,13)=2.376, R2=0.1545, slope not different from zero p=0.1472).*

3.4 Functional investigation of the pIC's role in fear extinction

The results from physiological recordings during acquisition and extinction of learned fear revealed that the pIC state-dependently processes fear-predictive cues, painful stimuli and that a large fraction of pIC neurons encodes states of high anxiety in large fractions of its neurons. As mentioned in the introduction, previous studies have already suggested that different subregions of the rodent insula may be involved in the acquisition and/or consolidation of learned fear [101–104]. However, to this date, no study has investigated the functional role of the pIC in the extinction of learned fear. Given the results from my fiber photometric and electrophysiogical recordings, I wanted to investigate whether pIC inhibition during extinction has an influence on the rate and/or effectiveness of extinction learning.

To characterize the functional implications of the pIC on fear extinction, I used optogenetics to broadly inhibit its activity.

3.4.1 Validation of optogenetic pIC inhibition using optrode recordings

As a first step, I aimed to validate optogenetic inhibition using optrode recordings in the pIC to simultaneously optogenetically inhibit and record neuronal activity. To this end, I infected excitatory pyramidal neurons of the pIC by injecting an AAV coding for Halorhodopsin under the CamKIIa promotor (AAV2/5-CaMKIIa-eNpHR3.0-eYFP) and implanted an optrode (a combination of an electrode bundle with an optic fiber) above the injection site (**Figure 19a)**. The pIC of three mice was inhibited using a series of five 40 s long, constant yellow light pulses, and a total of 17 single units was recorded as described above (see **2.6.2**). I found that photo-inhibition reliably reduced the firing rate of the majority of all recorded neurons from all mice (**Figure 19b, c and d)**. An immediate and stable decrease in firing rate could be observed upon light exposure in 12 out of 17 recorded neurons (71%, **Figure 19d**). Surprisingly, I also detected a small number of units where firing rate was significantly increased during light exposure (n=5 neurons, 29%), but the dynamics of this firing rate increase was slower, not stable, and smaller in absolute magnitude than for the inhibited neurons (**Figure 19d)**, possibly due to a disinhibition mechanism (see **Limitations and outlook).**

Taken together, optogenetic pIC silencing reliably reduced neuronal firing in a majority of pIC neurons and was thus an appropriate methods for the following functional investigation experiments.

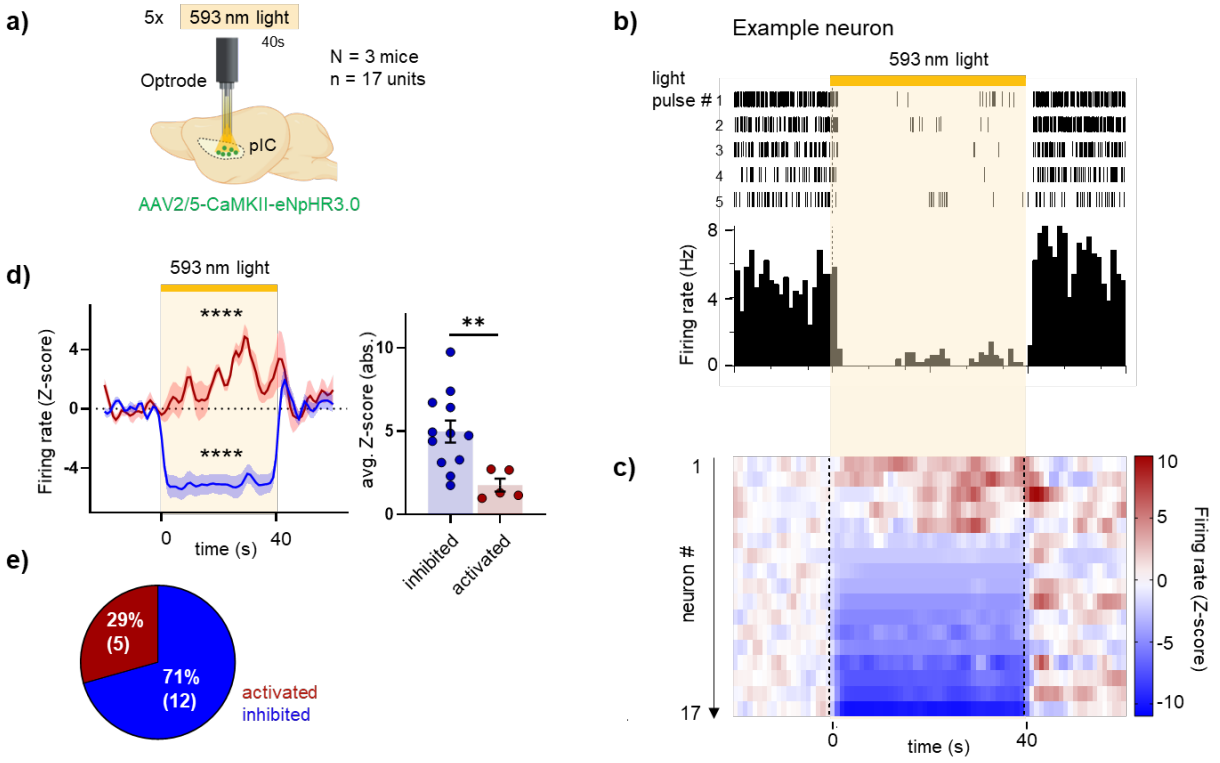


Figure 19: Validation of optogenetic inhibition of pIC neurons.

a) *Virus injection and optrode (16-wire electrode bundle fused to an optic fiber) placement above the pIC. Note that only excitatory cells were expressing the opsin. Light (593 nm, 12 mW at fiber tip) was delivered five times for 40 s. In total, n = 17 neurons from N = 3 mice were recorded.* ***b)*** *Raster plot and peri-event histogram of an example neuron aligned to the beginning of 40 s of light exposure (yellow shading; bin size = 1 s).* ***c)*** *Heatmap of z-scored neuronal activity of all recorded neurons (average across five light exposures). Vertical dashed lines indicate start and end of light delivery.* ***d)*** *Left: Average z-scored population response to the five light exposures (yellow shading). One-sample t tests with zero as theoretical mean revealed that during light exposure, a large population of neurons was significantly inhibited (blue, t=55.47, df = 39, p****<0.0001) and another, smaller population was significantly activated (red, t=8.385, df = 39, p****<0.0001) Note that the dynamic of the activated population was slower and less stable in comparison to the inhibited population. Right: The absolute average z-score of the inhibited neurons was significantly higher in comparison to the activated neurons (unpaired two-tailed t test, t=2.991, df=15, p**=0.0091). All data are plotted as mean ± SEM.* ***e)*** *A large fraction of all recorded neurons was significantly inhibited during light exposure (71%), whereas only a small population was (probably indirectly via a disinhibition mechanism) activated (29%).*

3.4.2 Optogenetics during fear extinction

I next aimed to optogenetically inhibit the pIC while animals were undergoing extinction. For these experiments, I used the same approach for viral injections as for the optrode recordings, except that the pIC was infected bilaterally and optic fibers were implanted above both injection sites ('eNpHR3.0'

group). As control condition, I used mice that were also bilaterally implanted with optic fibers, but injected with a viral vector expressing only the fluorophore (AAV2/5-CaMKIIa-eYFP, 'eYFP' group) (**Figure 20a)**.

3.4.2.1 Effects of pIC inhibition on extinction learning

Using the same FC/Ext paradigm as in **Figure 6**, I inhibited excitatory pIC neuron activity during CS+ presentations during extinction (**Figure 20b**). Although both groups equally expressed conditioned fear behavior to the CS+ (indicated by a similar time spent freezing to the CS+ during recall), eNpHR3.0 seemed to have a facilitated extinction learning (indicated by a faster reduction of CS+ evoked freezing with progressing extinction) in comparison to the eYFP controls (**Figure 20c**). Indeed, there was a significant difference between eYFP and eNpHR3.0 animals in CS+ evoked freeing during extinction late (**Figure 20d)**. Note that during extinction retrieval, both groups show similar low freezing levels**.** Taken together, these results suggests that pIC inhibition facilitated within session extinction learning, and that this extinction memory was stable as it could successfully be retrieved 24 hours later.

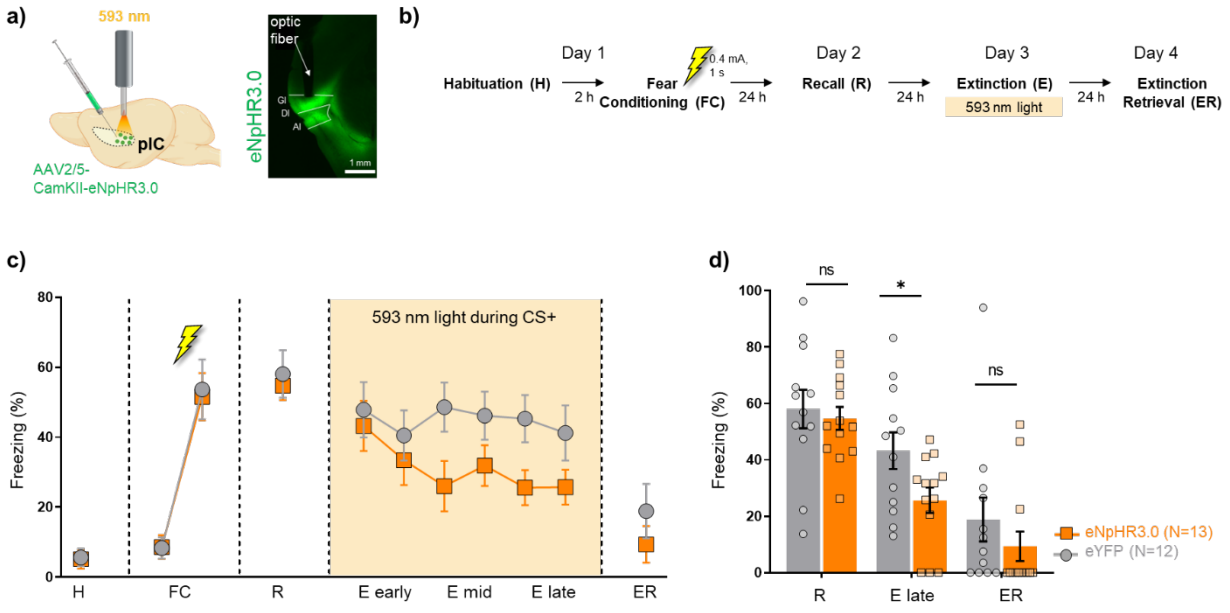


Figure 20: pIC silencing during extinction learning facilitates the extinction of a weak fear memory. ***a)*** *Left: Virus injection and optic fiber placement above the pIC. Right: Example of histological verification of virus expression and fiber placement. GI, granular IC; DI, dysgranular IC; AI, agranular IC.* ***b)*** *FC/Ext paradigm used (same as in Fig. 6). pIC was optogenetically inhibited during CS+ presentations during extinction (E) in eNpHR3.0:pIC animals.* ***c)*** *Although both groups equally expressed fear behavior to the CS+ (indicated by around 57% time spent freezing to the CS+ during recall (R)), animals with optogenetic pIC inhibition show a trend towards less freezing during extinction in*

comparison to the eYFP controls (N=13 eNpHR3.0, N=12 eYFP, two-way RM ANOVA of extinction, group (opsin) effect, F(1, 24)=2.678, p=0.1148; time effect, F(2.806, 67.34)=2.987, p<0.0403, group x time interaction F(5, 120)=1.036, p=0.3995).* ***d)*** *Comparison of average freezing during recall (R), extinction late (E late) and extinction retrieval (ER) revealed that even though both groups were freezing equally during recall, eNpHR3.0 animals showed an enhanced extinction learning in comparison to the eYFP controls; note that there was no difference in freezing during extinction retrieval between the groups (two-tailed unpaired t tests between groups, R: t=0.2549, df=23, p=0.8011, extinction late: t=2.120, df=23, p*=0.0450, ER: t=0.7953, df=23, p=0.4346). All data are plotted as mean ± SEM.*

However, the observed effects of pIC inhibition on extinction learning were rather weak, which could have been due to a floor effect, as both groups were not freezing more than around 50% at the beginning of extinction. I hypothesized that if initial freezing levels were higher, the extinction facilitating effects of pIC inhibition might become more obvious. It has previously been shown that by increasing the foot shock (US) intensity, it is possible to augment fear learning [226]. Therefore I decided to increase the strength of the US, but leaving all other parameters unchanged (**Figure 21b)**. To this end, a small, naïve group of animals underwent a 'strong FC' paradigm using a high intensity foot shock (0.6 mA and 2 s). Indeed, the initial amount of fear learning was increased and similar between both groups (around 70% of freezing to the CS+ during recall). Surprisingly, the effect of pIC inhibition on extinction learning was inverted in comparison to my initial results: eNpHR3.0 animals showed a clear impairment of extinction learning in comparison to the eYFP controls (**Figure 21c).** Analysis of CS+ evoked freezing during extinction late revealed that eNpHR3.0 mice were freezing significantly more in comparison to eYFP controls (**Figure 21d)**. Note that there is no significant difference in freezing during extinction retrieval between the groups (**Figure 21d)**. In summary, after fear conditioning with a high intensity US, pIC inhibition seemed to impair within session extinction learning.

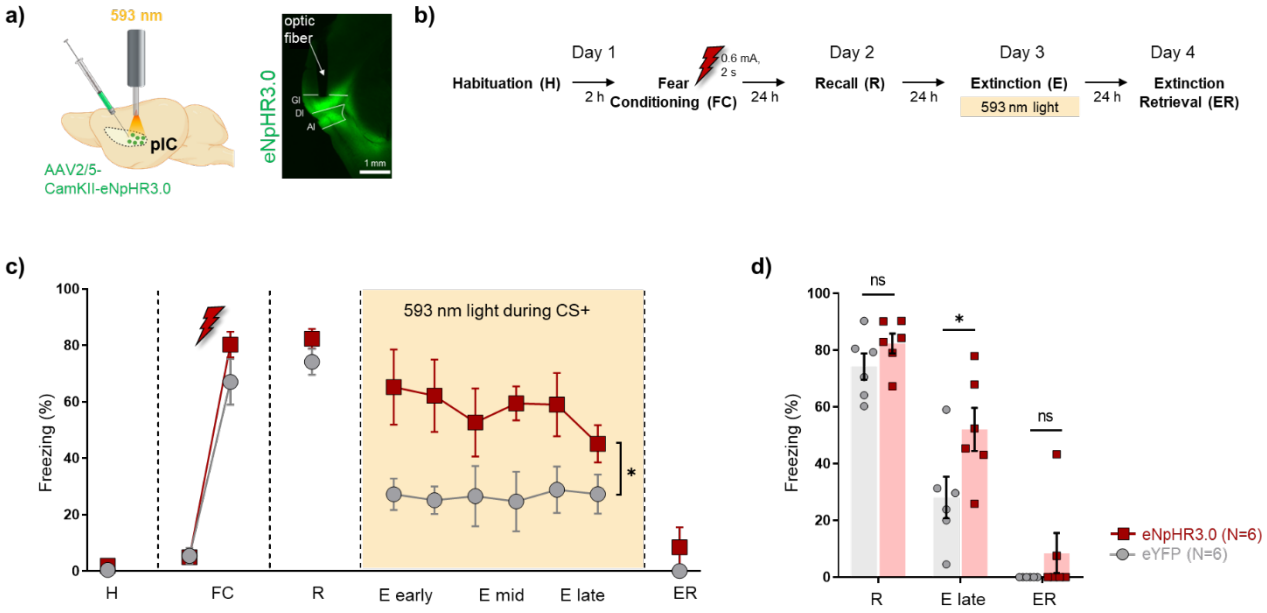


Figure 21: pIC silencing during extinction learning impairs the extinction of a strong fear memory
a) *Left: Virus injection and optic fiber placement above the pIC. Right: Example of histological verification of virus expression and fiber placement. GI, granular IC; DI, dysgranular IC; AI, agranular IC.* ***b)*** *Fear conditioning and extinction paradigm (similar as in Figure 20), but using a stronger US (0.6 mA for 2 s). pIC was again optogenetically inhibited during CS+ presentations during extinction (E) in eNpHR3.0 animals.* ***c)*** *Although both groups equally expressed conditioned fear behavior to the CS+ during recall (R), eNpHR3.0 animals showed significantly higher freezing during extinction in comparison to the eYFP controls (N=6 eNpHR3.0, N=6 eYFP, two-way RM ANOVA of extinction, group (opsin) effect, F (1, 10)=10.34, p*=0.0092; time effect, F(2.404,24.04)=0.4627, p=0.6698, group x time interaction F(5,50)=0.5419, p=07436).* ***d)*** *Comparison of average freezing during recall (R), extinction late (E late) and extinction retrieval (ER) revealed that eNpHR3.0 animals showed an impaired reduction in freezing to the CS+ in comparison to the eYFP controls during extinction late; note that there was no significant difference in freezing during recall and extinction retrieval between the groups (two-tailed unpaired t tests between groups, R: t=1.413, df=10, p=0.1882, extinction late: t=2.276, df=10, p*=0.0461, ER: t=1.197, df=10, p=0.2590). All data are plotted as mean ± SEM.*

3.4.2.2 pIC inhibition during extinction learning upon consecutive weak and strong FC

After having observed the paradoxical, bidirectional effects of pIC inhibition on extinction learning, I next asked whether I could replicate both results within the same animal and thus uncover whether the US intensity really is what determines the direction of the effect. To this end, I repeated the experiments with a naïve group of eNpHR3.0 animals and their respective eYFP controls using a consecutive FC/Ext paradigm. This consecutive paradigm consisted of a ‘weak’ (using a weak US 0.4 mA and 1 s), and a subsequent ’strong’ FC/Ext experiment (using a strong US 0.6 mA and 2 s), both with optogenetic pIC

inhibition during CS+ presentations during extinction (**Figure 22a and b)**. This consecutive FC/Ext protocol revealed that even though both groups were again equally fear conditioned after weak FC (as indicated by the same amount of freezing during recall, around 50%), pIC inhibition during extinction resulted in a mild facilitation of extinction learning that did not reach significance (**Figure 22 c)**. After strong FC, pIC inhibition caused an impairment in extinction learning in comparison to the eYFP control group, although it also did not reach significance (**Figure 22 c)**. Thus, there seemed to be an influence of US strength on the direction of the effect of pIC inhibition on extinction learning. Alternatively, I hypothesized that not the US strength itself but rather the FC-evoked anxiety levels might differ in intensity upon strong and weak FC procedures. Thus, it could be that the anxiety states of the individual mice might be driving the differential effects of pIC inhibition on FC performance. Indeed, a large number of studies has shown that the current emotional state of an individual strongly impacts learning and memory in general [227–230] and fear and extinction learning in particular [18,231–235]. Specifically, it has been suggested that optimal learning follows an inverted U-shaped curve, where low and high levels of arousal states generally impair performance and moderate levels facilitate it [228,236–238]. Emotional states can be modulated by varying US intensities [226,239,240] and subjects displaying a high fear state have been suggested to be more vulnerable to the influence of such stressors [241,242].

Therefore, I hypothesized that the observed bidirectional effects of pIC silencing on extinction learning might be due to individual differences in the animals' fear state. As a proxy for an animal's fear state, I thus analyzed how strong their initial fear memory after FC was by analyzing the average CS+ evoked freezing during recall. Interestingly, when I then correlated each mouse's fear state to its extinction performance (% freezing during extinction late), I found a strong positive correlation for eNpHR3.0 animals, both for extinction after weak and after strong FC, whereas these correlations for eYFP controls were much weaker **Figure 22d and e**). This suggests that when pIC activity was dampened during extinction learning, the internal fear state influenced extinction performance, whereas this influence was weakened when pIC activity was left undisturbed. Thus, it appears that under physiological conditions, the IC maintains fear and prevents too rapid fear extinction when animals are exhibiting low levels of arousal, whereas on the other hand, it alleviates extinction learning when animals are very aroused. This function could thus be adaptive in ensuring optimal fear levels to support best learning performances.

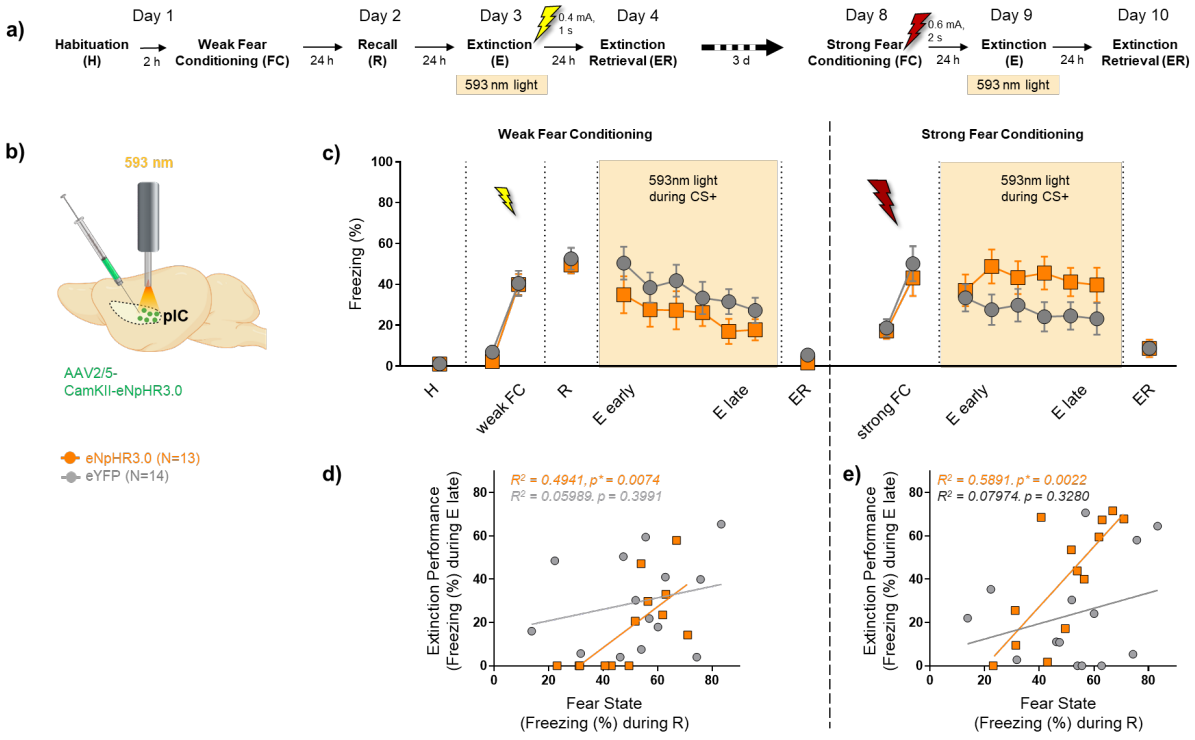


Figure 22: Consecutive weak and strong FC in the same animals has a bidirectional effect of pIC inhibition on extinction learning.

a) *Consecutive fear conditioning and extinction paradigm. Both eNpHR3.0 and eYFP control groups first underwent a weak FC/Ext paradigm using a normal, mild US strength (0.4 mA and 1 s duration, 'weak FC') with optogenetic pIC inhibition during CS+ during extinction (E). After a pause of 3 days, animals underwent a second round of FC/Ext using a high intensity US (0.6 mA and 2 s duration, 'strong FC'), again with optogenetic pIC inhibition during CS+ during extinction.* ***b)*** *Virus injection and optic fiber placement above the pIC.* ***c)*** *Freezing behavior of both groups across weak and strong FC and extinction plotted as average of % time freezing to two CS+ (FC, extinction, extinction retrieval) or four CS+ (habituation and recall). After weak FC, the eNpHR3.0 group shows a trend towards facilitated extinction learning in comparison to eYFP controls, whereas the same eNpHR3.0 animals show a trend towards impaired extinction learning after strong FC (N=13 eNpHR3.0, N=14 eYFP, two-way RM ANOVA of extinction after weak FC: group (opsin) effect, F (1, 25)=1.539, p=0.2263; time effect, F(5,125)=0.4627, p****<0.0001, group x time interaction F(5,125)=0.5427, p=0.7436; two-way RM ANOVA of extinction after strong FC: group (opsin) effect, F (1, 25)=27.60, p=0.1320; time effect, F(5,125)=1.105, p=0.3611, group x time interaction F(5,125)=1.987, p=0.0851). Data are plotted as mean ± SEM.* ***d)*** *Correlation between 'fear state' (% time spent freezing during recall) and 'extinction performance' (% time spent freezing during extinction late) for weak FC. Linear regression reveals a strong positive correlation in eNpHR3.0 animals (F(1,11)=10.74, R2=0.4941, slope significantly non-zero, p**=0.0074) whereas in eYFP controls, the correlation is much weaker (F(1,12)=0.7645, R2=0.05989, slope not different from zero, p=0.3991; comparison of both slopes: ANCOVA, F(1,23) = 2.312, P=0.1420).* ***e)*** *Correlation between 'fear state' and 'extinction performance' for strong FC also reveals a strong positive correlation in eNpHR3.0 animals(F(1,11)=15.77, R2=0.5891, slope significantly non-zero, p**=0.0022), whereas the correlation is much weaker in eYFP controls (F(1,12)=1.040, R2=0.07974, slope not different from zero, p=0.3280; comparison of both slopes: ANCOVA, F(1,23) = 3.889, P=0.0607).*

To corroborate these findings, I also applied the same correlation analysis to the data from both previous optogenetic experiments described above (see **Figure 20** and **Figure 21**). Interestingly, I could observe a similarly strong positive correlation between fear state and extinction performance in eNpHR3.0 animals, but not in the eYFP controls (*Supplementary figure 1*). Additionally, there was also no correlation between fear state and extinction performance in animals from **Figure 6** where pIC activity was left undisturbed (**Supplementary figure 4**).

Notably, in the data from consecutive weak and strong FC, there was a large and continuous spread in CS+ evoked freezing during recall (i.e. fear state) among individual animals, ranging from just above 10% to more than 80% time spent freezing to the CS+ (**Figure 23a)**. To classify animals into two distinct groups ('low fear state' and 'high fear state' group), I divided them using a threshold of 50 % CS+ freezing during recall (**Figure 23a)**, yielding a low fear state group that consisted of N=6 eNpHR3.0-low fear and N=5 eYFP-low fear mice, and a high fear state group that consisted of N=7 eNpHR3.0-high fear and N=9 eYFP-high fear mice.

I then re-analyzed the data from consecutive FC/Ext by splitting the data from high and low fear state groups. I found that for high fear state animals, the extinction impairment after strong FC in eNpHR3.0 animals became more pronounced (although the group difference across the whole extinction session was only close to significant), whereas there was no difference in the freezing during extinction between eNpHR3.0 and eYFP animals after weak FC (**Figure 23b, top**). Accordingly, the extinction facilitation effect after weak FC became more evident in eNpHR3.0-low fear animals in comparison to eYFP-low fear animals (although also here, the group difference across the whole extinction session was only close to significant), whereas the groups did not differ in freezing during extinction after strong FC (**Figure 23b, bottom**). Quantitative analysis of extinction performance confirmed a significantly lower freezing during extinction late in eNpHR3.0-low fear mice than in eYFP-low fear mice after weak FC, and a significantly higher freezing during extinction late in eNpHR3.0-high fear mice than in eYFP-high fear mice after strong FC (**Figure 23c**). In contrast, all high fear state animals show the same extinction performance after weak FC, and all low fear state animals show the same extinction performance after strong FC (**Figure 23c**). Importantly, regardless of fear state or conditioning strength, all eYFP controls show the same rate of extinction learning, indicated by no group differences across the whole extinction session (**Supplementary figure 5a)** and by no significant differences in extinction performances (**Supplementary figure 5c)**. On the other hand, eNpHR3.0 animals show a large difference in extinction learning between the low and high fear state group regardless of conditioning strength, indicated by significant group differences across both extinction sessions (**Supplementary figure 5b)**, and by significant differences in extinction performances (**Supplementary figure 5d)**.

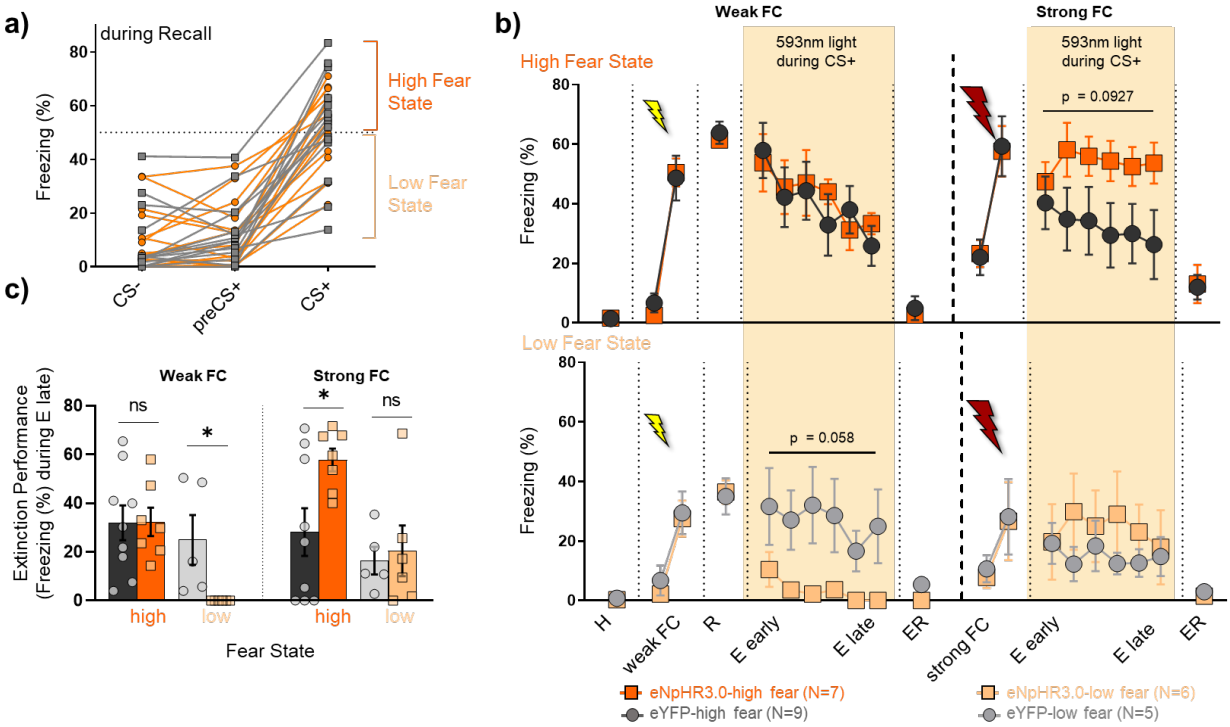


Figure 23: The internal fear state predicts the influence of pIC inhibition on extinction learning.
a) *Average % time spent freezing to all four CS-, preCS+ (30 s preceding each CS+), and CS+ during recall. Animals were separated into 'high fear state' and 'low fear state' groups according to their % freezing to CS+ (threshold at 50%).* ***b)*** *Same data as in Fig. 17c), but split into high (top) and low (bottom) fear state groups. In the high fear state group, eNpHR3.0 animals show a trend towards impaired extinction in comparison to the eYFP controls, but only after experiencing FC using a strong US (N=7 eNpHR3.0-high fear, N=9 eYFP-high fear, two-way RM ANOVA of extinction after strong FC: group (opsin) effect, F (1, 15)=3.225, p=0.0927; time effect, F(5,75)=0.6284, p=0.6786, group x time interaction F(5,75)=1.306, p=0.2707). In the low fear state group, however, eNpHR3.0 animals show a trend towards a rapid facilitation of extinction learning, but only after FC using a weak US (N=6 eNpHR3.0-low fear, N=5 eYFP-low fear, two-way RM ANOVA of extinction after weak FC: group (opsin) effect, F (1, 10)=4.601, p=0.0575; time effect, F(5,50)=3.332, p*=0.0113, group x time interaction F(5,50)=0.8744, p=0.5051). Data are plotted as mean ± SEM.* ***c)*** *Comparison of extinction performance (average % freezing during extinction late) in high and low fear state animals after weak and strong FC. Unpaired two-tailed t tests reveal extinction learning impairment in eNpHR3.0-high fear animals in comparison to eYFP-high fear animals after strong FC (t=2.475, df=14, p*=0.0267) and extinction learning facilitation in eNpHR3.0-low fear animals in comparison to eYFP-low fear animals after weak FC (t=2.698, df=9, p*=0.0245). Note that there is no difference in extinction performance between eNpHR3.0 and eYFP animals when high fear state animals experienced the weak FC (t=0.03984, df=14, p=0.9688), or when low fear animals experienced the strong FC (t=0.3206, df=9, p=0.7558). Data are plotted as mean ± SEM.*

In summary, the results described above indicate that under physiological circumstances, the pIC actively provides homeostatic regulation that ensures optimal extinction learning performance regardless of the emotional state. However, when the pIC is not engaged during extinction learning, this mechanism is

disrupted. Thus, a role of the pIC in extinction learning might be to maintain learning performances despite differences in emotional states.

3.4.2.3 pIC inhibition during shortened extinction learning

So far, my results have revealed within session alterations in extinction. However, extinction success is generally defined by the formation of lasting extinction memory [17]. In my studies, I can assess this memory formation through extinction retrieval 24 hours after extinction. However, in the results described above, I did not observe a difference in CS+ evoked freezing between eNpHR3.0 and eYFP control animals during extinction retrieval (see **Figure 20, Figure 21,** and **Figure 22**). This is most likely explained by a floor effect, since all animals, regardless of treatment or fear state, were given enough extinction training to completely extinguish and thus exhibited successfully consolidated fear extinction memory before entering the extinction retrieval session. In order to test whether pIC manipulations truly affected long-term extinction memory formation, I next tested a naïve group of eNpHR3.0 animals and their respective eYFP controls in a shortened FC/Ext paradigm, using only two CS+ during recall and eight CS+ during extinction, but leaving all other parameters unchanged **(Figure 24b**).

Indeed, correlation of fear state (i.e. freezing during CS+ in recall) and extinction retrieval performance (i.e. freezing during CS+ in extinction retrieval) showed a very similar pattern as described above, where in eNpHR3.0 animals the fear state could predict the extinction retrieval performance, but not in eYFP controls (**Figure 24c**). To better illustrate the fear behavior across the FC/Ext session, I divided the animals into low and high fear state groups as described above (50% freezing to the recall CS+) (**Figure 24d**). Nnote that in this cohort of mice, the high fear group was much larger (N=16 eNpHR3.0-high fear and N=13 eYFP-high fear) than the low fear group (N=5 eNpHR3.0-low fear and N=3 eYFP-low fear). Despite the fact that all animals from the high fear group showed no difference in freezing during extinction (**Figure 24e, top**), eNpHR3.0 animals exhibited significantly higher freezing during extinction retrieval in comparison to eYFP-high fear controls, indicating an impairment in extinction memory maintenance (**Figure 24f, left**). In contrast, the eNpHR3.0-low fear animals already showed a non-significant trend towards facilitated within session extinction (**Figure 24e, bottom**) and a (close to significant) trend to lower freezing during extinction retrieval, hinting towards a facilitation of extinction memory maintenance (**Figure 24f, right)**. Importantly, I again observed that eYFP-high fear and eYFP-low fear animals exhibited equal capabilities to retrieve extinction memory (**Supplementary figure 6a** and **c, left**), whereas eNpHR3.0-low fear and eNpHR3.0-high fear showed a significant difference in extinction memory retrieval (**Supplementary figure 6b** and **c, right**).

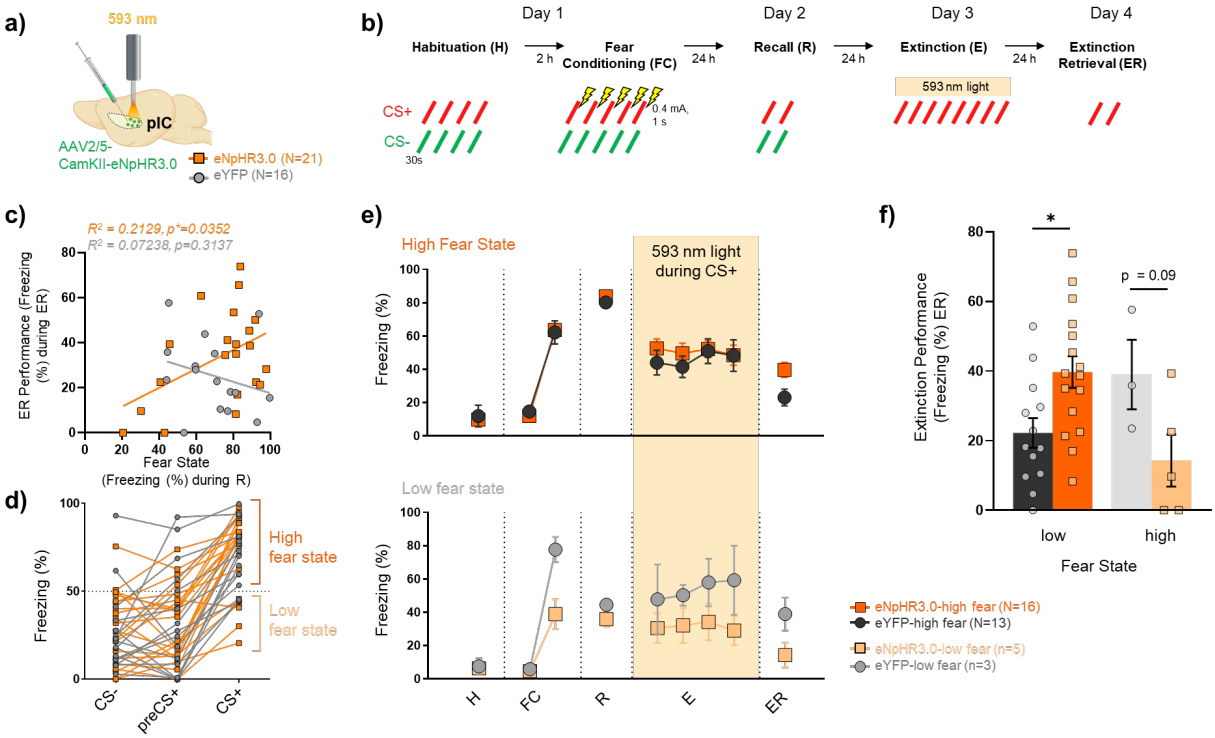


Figure 24: pIC silencing during shortened extinction learning has state-dependent bidirectional effects on extinction retrieval.

a) *Virus injection and optic fiber placement above the pIC.* ***b)*** *Shortened fear conditioning and extinction paradigm used: recall (R) consisted of only two CS+ and CS- presentations each and extinction (E) consisted of 8 CS+ in total. pIC was again optogenetically inhibited in eNpHR3.0:pIC animals during CS+ presentations during extinction. N=21 eNpHR3.0 and N=16 eYFP.* ***c)*** *Correlation between fear state (average % freezing to both CS+ during recall) and extinction retrieval (ER) performance (average % freezing to both CS+ during recall). Linear regression reveals that there is a strong positive correlation between fear state an extinction retrieval performance in eNpHR3.0 animals (F(1,19)=5.141, R2=0.2129, slope significantly different from zero, p*=0.0352), whereas eYFP controls show only a weak correlation (F(1,14)=1.092, R2=0.07238, slope not different from zero, p=0.3137).* ***d)*** *Average % time spent freezing to both CS-, preCS+ (30 s preceding each CS+), and CS+ during recall. Animals were separated into 'high fear state' and 'low fear state' groups according to their % freezing to CS+ (cutoff at 50%).* ***e)*** *Freezing behavior across the shortened FC and extinction experiment of both groups plotted as average of % time freezing to four CS+ (habituation) or two CS+ (FC, recall, extinction, extinction retrieval). N=16 eNpHR3.0-high fear, N=13 eYFP-high fear, N=5 eNpHR3.0-low fear, N=3 eYFP-low fear. Data are plotted as mean ± SEM.* ***f)*** *Comparison of extinction retrieval performance after shortened extinction between high and low fear state animals. Unpaired two-tailed t tests reveal extinction retrieval is impaired in eNpHR3.0-high fear animals in comparison to eYFP-high fear animals (t=2.765, df=27, p*=0.0101). In contrast, eNpHR3.0-low fear animals show a trend towards extinction retrieval facilitation in comparison to eYFP-low fear animals (t=1.996, df=6, p=0.0930). Data are plotted as mean ± SEM.*

In conclusion, these results confirm that the pIC truly affects extinction performance, both within and beyond the extinction learning session. Thus, the effects of pIC inhibition on extinction performance are not simply due to anxiogenic or anxiolytic effects during the extinction sessions itself.

3.5 Manipulation of interoceptive processes during extinction learning

Taken together, my results so far established that the pIC encodes and processes anxiety-states and that it exerts a state-dependent influence on fear extinction learning. States of heightened anxiety are characterized by strong bodily changes, such as changes in heart rate, breathing, sweating, or intestinal motility. Indeed, famous emotion theories postulate that some emotion states may arise or are at least strongly influenced by interoception (see [26,27,243] for comprehensive reviews). As mentioned in the introduction, the insula is the main brain cortical region involved in interoceptive processing, and bodily signals are transmitted to the brain mainly through the vagus nerve. Interestingly, several studies on rats have already shown that when the vagus nerve is stimulated during extinction, their learning performance is enhanced [206,208,209,212,213].

Therefore, I next wanted to test the possibility that interoceptive signals may influence fear extinction performance and may be processed by the insular cortex.

3.5.1 Effects of VNS on extinction learning

To test the influence of interoceptive processes on insular activity as well as fear extinction learning, I applied vagus nerve stimulation (VNS) during extinction (**Figure 25a)**. VNS was paired with CS+ presentations since in FP as well as single unit recordings, I observed strong CS+ evoked pIC responses during heightened anxiety states (**Figure 8** and **Figure 14**).

The behavioral protocol for FC/Ext was similar to the protocol I used in previous experiments, but had to be adapted as previous studies have reported that VNS effects on extinction were progressively developing across several days of VNS exposure [211]. Thus, animals were undergoing three consecutive days of extinction, where the first two CS+ presentations were not paired with VNS, followed by ten CS+ presentations that were paired with VNS (31s duration (starting 1 s before CS+ onset), 0.4 mA, 30 Hz, 500µs pulse width, **Figure 25b**). Accordingly, the first two unpaired CS+ of the extinction 1 served as fear memory recall (and thus also as the fear state of the animal), and the unpaired CS+ of extinction 2 and 3 served as test for extinction retrieval. As controls ('Sham'), I used animals that underwent the same surgery, but did never receive any stimulation of the vagus nerve.

Results from FC/Ext paired with VNS showed that VNS has a state-dependent bidirectional effect on extinction learning (**Figure 25c to e)**. The fear state (% freezing during recall) of VNS animals was highly correlated to their extinction performance (% freezing during extinction retrieval 1), whereas this

correlation was not apparent in Sham controls (**Figure 25c).** Therefore, I separated the animals according to their fear state using the same threshold as described above (50% freezing to the recall CS+), resulting in a high and a low fear state group (**Figure 25d)**. When comparing the CS+ evoked freezing across the whole experiment between the groups (**Figure 25e)** it became even more obvious that the VNS high fear animals showed an significantly impaired extinction learning compared to the Sham high fear animals, and that the VNS low fear animals had a significant enhancement in extinction learning in comparison to the Sham low fear controls. Importantly, Sham control animals extinguish at an almost identical rate throughout extinction learning (extinction 1 to extinction 3) irrespective of their fear state, whereas high and low fear VNS animals show significantly different freezing levels during extinction learning (**Supplementary Figure 7**).

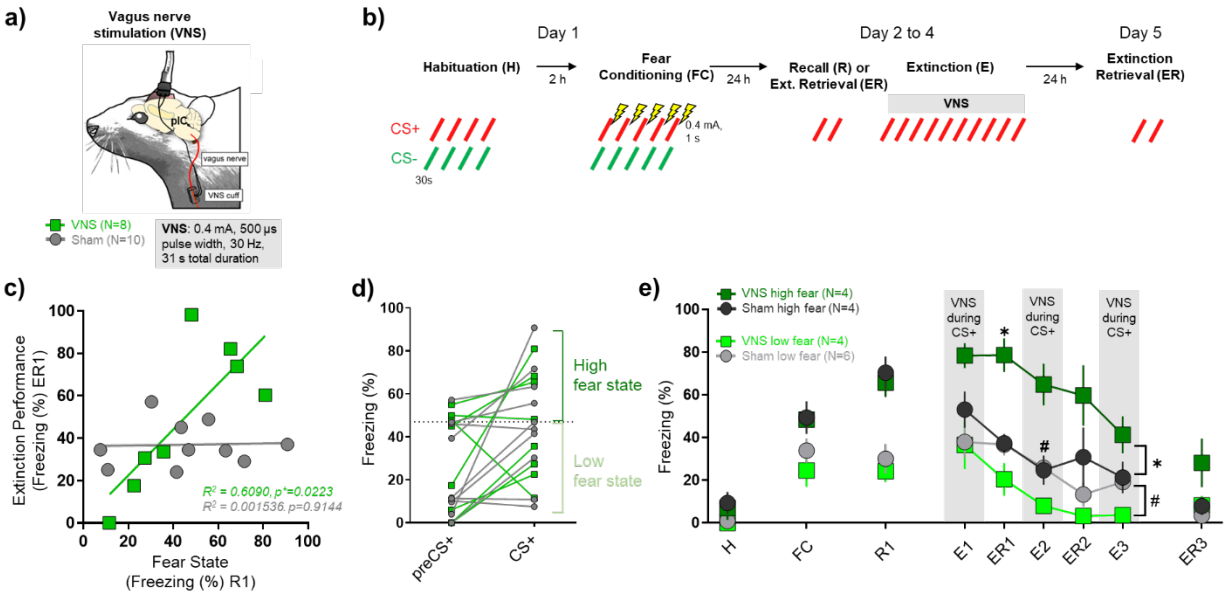


Figure 25: Vagus nerve stimulation (VNS) renders fear extinction state-dependent.
a) *VNS cuff implantation approach and VNS parameters (0.4 mA, 500 µs pulse width, 30 Hz, 31 s total duration). N=8 VNS animals, N=10 Sham control animals.* ***b)*** *Fear conditioning and extinction paradigm used was similar to the paradigms used before with the following alterations: recall (R) and extinction (E) happened in the same session, and consisted of two CS+ presentations for recall and ten CS+ presentation for extinction. During extinction, the CS+ presentations were coupled to VNS. The same combination of recall and extinction was repeated on days 3 and 4 of the experiment and extinction retrieval was tested on day 5. Note that the recall and extinction session are numbered consecutively.* ***c)*** *Correlation between fear state (average % freezing to both CS+ during recall 1) and extinction performance (average % freezing to both CS+ during extinction retrieval 1). Linear regression reveals that there is a strong positive correlation between fear state an extinction performance in VNS animals ($F(1,6)=9.345$, $R^2=0.6090$, slope significantly different from zero, $p^*=0.0223$), whereas Sham controls show only a weak correlation ($F(1,8)=0.01231$, $R^2=0.001536$, slope not different from zero, $p=0.9144$).* ***d)*** *Average % time spent freezing to both preCS+ (30 s preceding each CS+), and CS+ during recall. Animals were separated into 'high fear state' and 'low fear state' groups according to their % freezing*

to CS+ (cutoff at 50%). ***e)*** *Freezing behavior across the FC/Ext experiment of both low and high fear groups plotted as average of % time freezing to all CS+ presentation for each stage. Two-way RM ANOVA of freezing across all extinction days (extinction 1 to extinction 3) revealed that VNS high fear animals show a strong extinction learning impairment in comparison to the Sham control group (group (stimulation) effect, F (1, 6)=7.515, p*=0.0337; time effect, F(1.597,9.585)=13.37, p**=0.0024, group x time interaction F(4,24)=1.652, p=0.1939; Bonferroni post-hoc VNS vs Sham, p*=0.0363 for ER1). In contrast, low fear VNS animals show a strong facilitation in extinction learning in comparison to the Sham controls group (two-way RM ANOVA, group (stimulation) effect, F (1, 8)=14.42, $p^{\#}$=0.0053; time effect, F(2.384,19.07)=7.585, $p^{\#\#}$=0.0026, group x time interaction F(4,32)=0.5823, p=0.6777; Bonferroni post-hoc VNS vs Sham, $p^{\#}$=0.0320 for E2). Data are plotted as mean ± SEM.*

In conclusion, I could show that interoceptive signaling via the vagus nerve indeed influences extinction performance, and that VNS and pIC inhibition have similar state-dependent and bidirectional effects on extinction learning. Given the similarity of the effects and the suggested role of the pIC as the major cortical recipient of interoceptive information, I hypothesized that VNS may have a direct influence on pIC activity which influences its role in fear extinction.

3.5.2 Effects of VNS on pIC activity

To test whether VNS modulates pIC activity, I combined VNS with fiber photometry recordings in pIC while animals were exposed to several rounds of vagus nerve stimulations in a neutral environment ('VNS only'). Afterwards, the same animals were undergoing the FC/Ext paradigm where CS+ was paired with VNS during extinction to analyze potential correlations between VNS evoked responses in pIC and extinction learning.

To this end, I implanted a naïve cohort of 13 mice with stimulation cuffs around the left vagus nerve, infected ipsilatera pIC with an AAV coding for a calcium indicator under the CaMKIIa promotor (AAV9.Camk2a.GCaMP6s.WPRE.SV40) and implanted an optic fiber above the virus injection site (**Figure 26a top**). The 'VNS only' experiment was conducted by placing the animals into a neutral environment and applying ten bouts of VNS (with the same parameters that were used in the previous experiment) and recorded from excitatory pIC neurons throughout the whole session (**Figure 26a, bottom**). I found that in some animals, there was a very strong and significant increase in pIC activity upon each VNS bout ('VNS responders', N=5) (**Figure 26b and c)**. However, there was a larger fraction of animals where I could not observe a significant VNS evoked response ('non-responders', N=8) (**Figure 26c**). When I then tested both groups (VNS responders and non-responders) in the FC/Ext paradigm, I found that both groups exhibit similar levels of freezing during recall, indicating that they acquire conditioned fear responses equally. Remarkably, VNS during extinction causes a rapid and significant decrease in CS+ evoked freezing in the VNS responder group, already after one day of extinction learning,

as indicated by a significantly lower % freezing in VNS-responders in comparison to non-responders during extinction retrieval 1 (**Figure 26d**). Even though there was no significant group difference between VNS-responders and non-responders during the extinction days (extinction 1 to extinction 2) (which could also be due to rather small N numbers here), there was a significant time as well as group x time interaction effect, indicating that the VNS responders extinguished faster than the non-responders.

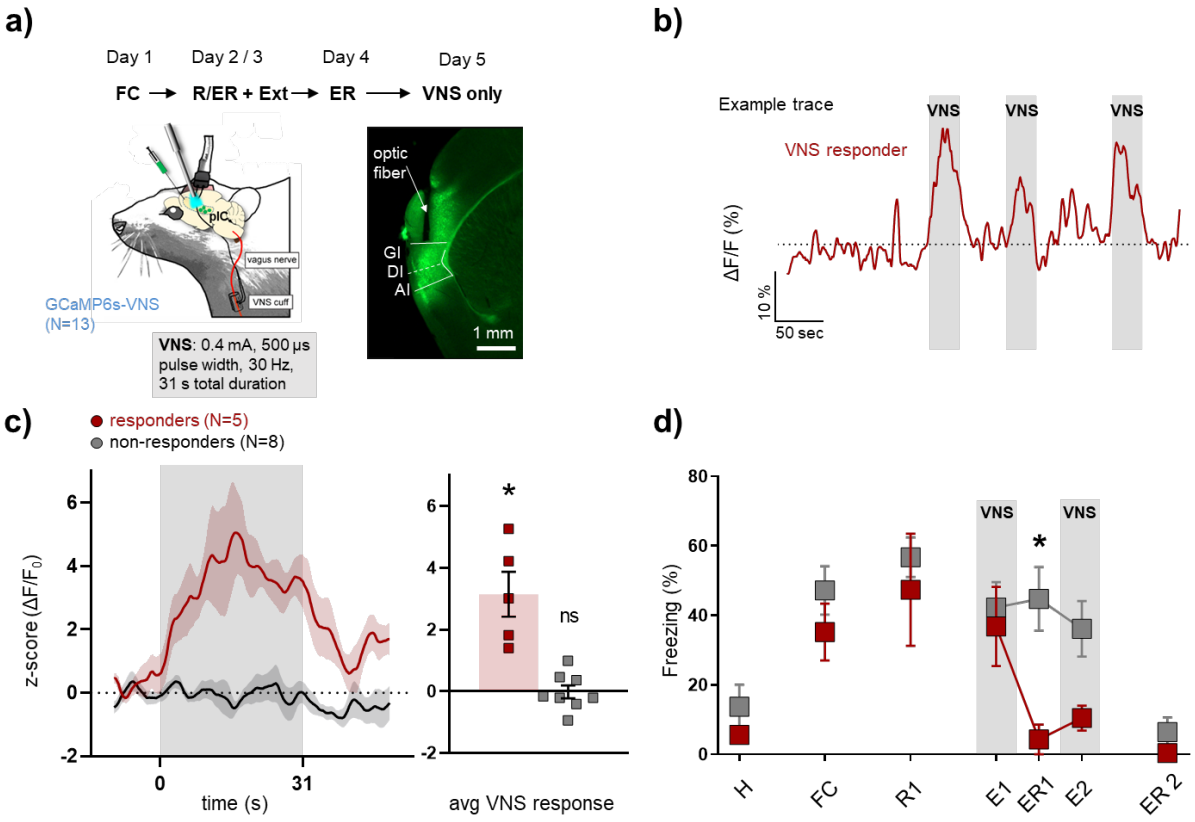


Figure 26: VNS modulates pIC activity and influences extinction learning in a subset of animals.
a) *Top left: VNS cuff implantation, virus injection and fiber implantation approach and VNS parameters (0.4 mA, 500 µs pulse width, 30 Hz, 31 s total duration). Top right: Histological verification of virus expression and fiber placement. GI, granular IC; DI, dysgranular IC; AI, agranular IC. Bottom: Experimental time line. N=13 GCaMP6s-VNS animals.* ***b)*** *and* ***c)*** *Data from 'VNS only' recordings on day 1 of the experiment.* ***b)*** *Example GCaMP6s trace from a 'VNS responder' mouse, grey shadings indicate VNS timings.* ***c)*** *Left: Average traces of pIC excitatory neuron activity during 31 s of VNS (grey shading). Traces are showing averaged z-scored neuronal responses to all ten VNS presentation of 'VNS responders' (red, N=5) and 'non-responders' (black, N=8). Right: Average z-score during VNS of each animal. Two-sided one sample t tests with zero as theoretical mean revealed significant pIC excitatory neuron activation in VNS responders (t=4.351, df=4, p*=0.0121; non-responders: t=0.07757, df=7, p=0.9403).* ***d)*** *Freezing behavior to the CS+ during the FC/Ext paradigm (days 2 to 5 of the experiment). VNS responders show a strong extinction facilitation that develops over time and is apparent from extinction retrieval 1 on (two-way RM ANOVA of extinction days (extinction 1 (E1) to extinction 2 (E2)), group (stimulation) effect, F (1, 12)=3.135, p=0.1020; time effect, F(2, 24)=6.2320, p**=0.0066, group*

x time interaction F(2,24)=5.590, p=0.0102; Bonferroni post-hoc VNS responders vs non-responders, p*=0.0238 for ER 1). Data are plotted as mean ± SEM.*

4 DISCUSSION

In my PhD thesis, I revealed that the pIC processes fear evoking cues in a state-dependent manner and encodes sustained states of anxiety in populations of single pIC neurons. I further discovered a seemingly paradoxical, bidirectional influence of the pIC on extinction learning. I could reveal that the opposing influence of pIC silencing on extinction performance was due to differences in the internal fear state of the manipulated animals, thus suggesting a role for the intact IC in homeostatic regulation of fear extinction performance across varying levels of anxiety.

Further, I established preliminary evidence that the state-coding of the pIC may be largely influenced or driven by interoceptive signals transmitted to the brain via the 10th cranial nerve (the vagus nerve).

While human and rodent studies have suggested a role for the IC in processing fear evoking cues and anxiety states [88,89,92,93] as well as in processing interoceptive signals [23,27], my study is the first to provide a first mechanistic insight into how bodily signals mediated via the vagus nerve can influence pIC activity to cause differential effects on fear extinction.

4.1 pIC inhibition exhibits a state-dependent bidirectional effect on extinction learning

In order to elucidate the role of the pIC in extinction learning, I here optogenetically inhibited the pIC during CS+ presentations in the extinction session. I found a surprising, bidirectional effect on extinction learning that was dependent on the animal's fear state: in high fear state animals, pIC silencing lead to an impairment in extinction learning or extinction retrieval, respectively, in comparison to eYFP controls, whereas in low fear state animals, a facilitation was observed.

Interestingly, when animals were undergoing 'weak' FC/Ext (using only a mild US) (**Figure 20**), the dominating effect of pIC silencing was an extinction facilitation, whereas in the cohort that experienced 'strong' FC/Ext (using a US with higher duration and intensity) (**Figure 21**), the dominating effect was an extinction impairment. When the same animals were consecutively undergoing first weak and then strongFC/Ext, the bidirectionality of the effect became even more pronounced (**Figure 22**).

Importantly, I found that in eNpHR3.0 animals, the internal fear state was able to predict the direction of the effect, whereas this was not possible in eYFP controls. As a proxy for the internal fear state, I used the amount of fear behavior expressed during CS+ presentations in the recall session, which reflects how strong the animal was initially fear conditioned. Interestingly, there is a strong inter-individual difference in fear state within the same cohort of mice, despite the fact that they all received the same US intensity during FC. Thus, I hypothesized that the fear state (and thus also the direction of the effect of optogenetic pIC inhibition) is not only determined by the strength of the US. Rather, the fear state is a combination of an animal's "trait anxiety" and the intensity of the aversive experience. Trait anxiety is regarded as a stable biological predisposition [244] that reflects an individual's general disposition to experience anxiety-relevant feelings or to show anxiety evoked behaviors [245]. Further, high trait anxiety individuals have the tendency to respond fearfully to a wide variety of unspecific aversive stimuli [246]. It has been shown that also rodents display such anxiety trait related behaviors, and they can selectively be bread for high and low anxiety traits [247]. As there probably is a continuous spectrum of trait anxiety levels, there is also a continuum in the observed results from this study: For instance, a low trait anxiety mouse receiving a mild foot shock will show an extinction facilitation effect, whereas a medium trait anxiety mouse receiving the same foot shock might not show an effect at all, etc. To test this hypothesis, in subsequent experiments one would need to screen for individual anxiety-trait levels before FC/Ext experiments, for example using an innate fear test (EPM or OFT) and analyze if trait anxiety level in combination with foot shock strength can predict the fear state, and thus the direction of the effect of pIC inhibition on fear extinction.

A very interesting observation from this study was that eYFP control animals were showing very similar extinction performances and extinction retrieval performances, regardless of their fear state (**Supplementary figure 5, Supplementary figure 6,** and **Supplementary figure 7)**. Whereas other studies in rodents have shown that an animal's fear and stress levels have a strong influence on extinction performance [139,248], in my study, both high and low fear state eYFP animals extinguished at a similar rate and could also retrieve extinction memory equally. Thus, it seems that an important role of the pIC is to maintain optimal extinction learning performance, regardless of the actual fear state. When the contribution of the pIC is removed, this homeostatic control mechanism seems to be impaired, and the fear state of an animal determines its extinction performance. How this homeostatic mechanism is mechanistically implemented remains a vital subject for future studies.

It is also important to mention that the observed extinction facilitation or extinction impairment effects could also be explained by 1) an acute influence of pIC activity on freezing/motor behavior or 2) and acute anxiolytic or anxiogenic effect of pIC inhibition, respectively. However, I can fully exclude option 1) as in a previous study from our lab where I was a co-author, we have found no effect of pIC inhibition or activation on distance travelled in an EPM or OFT test [35]. On the other hand, in the same study, we

have shown that global optogenetic pIC inhibition indeed influences innate anxiety behaviors in an EPM or OFT [35]. However, we could only observe anxiolytic, but never an anxiogenic effects. (Interestingly, anxiogenic effects were only detected with optogenetic pIC stimulation.) It cannot be excluded that anxiolysis caused by pIC inhibition might have at least supported the extinction facilitation effects. Research on the treatment of human anxiety disorders shows a growing interest in the application of anxiolytic drugs for the augmentation of fear extinction learning, but most of these approaches include systemic pharmacological treatments that might have severe side effects[144,249]. Thus, targeted manipulations of IC activity (for example using transcranial direct current stimulation (tDCS)), might be a relevant adjuvant for enhancing exposure-based therapies in the future. Nevertheless, one should be extremely careful with this translational approach, since I have also shown that the effects of insula manipulations are state-dependent and could cause unwanted extinction impairment effects in individuals exhibiting a more fearful emotional state. Thus, more research regarding individual differences has to be conducted to make insula manipulations a valuable and safe translational tool.

Taken together, I suggest that the bidirectional influence of pIC inhibition of extinction learning are caused by a more complex, state-dependent and learning-related mechanism, and cannot solely be explained by an acute influence on fear behavior.

4.2 Interoceptive processes influence pIC activity and extinction learning

One of the most striking findings I made in this study was that stimulation of the vagus nerve had the same state-dependent and bidirectional effects on extinction learning as optogenetic pIC silencing (**Figure 25**). Previous research on rats has repeatedly shown a rapid remission of conditioned fear responses when VNS was paired with CS+ presentations during extinction [208,210,213,250]. In my study, I could replicate this effect, but only in a subgroup of mice that exhibited a low internal fear state. However, in the high fear state group, VNS caused impaired extinction, an effect that has never been reported before. Additionally, it has even been reported that in a rat model of PTSD that shows robust extinction impairments, VNS was able to completely reverse these impairments [211]. On the contrary, in my study, mice exhibiting a high fear state showed a robust fear extinction impairment upon VNS treatment.

Numerous VNS studies in humans and rodents indicate that stimulation intensity, frequency, number of stimulations, timing of stimulations and inter-stimulation intervals (ISI) are critically influencing the effect of VNS on synaptic plasticity [251–253], memory formation [254,255] as well as extinction of conditioned fear [250]. Interestingly, several studies have also suggested that VNS effects may follow an inverted-U

profile, where better effects are observed in moderate than at low or high intensities [253,256,257]. In high fear state animals, the vagus nerve might already show an elevated activity, as it provides information about peripheral arousal to the brain. Additional stimulation might then cause a shift towards negative effects of VNS on extinction learning. Along the same line, vagus activity of low fear state animals might be comparatively low, but shifted towards moderate levels upon VNS, thus causing an improvement in extinction learning.

While many research laboratories actively sought after the neuronal mechanisms of VNS, we currently still lack a mechanistic understanding of the efficacy of VNS. One suggestion is that VNS works via stress-evoked memory enhancing effects. Several studies in humans and rodents indicate that a strong sympathetic stress response during exposure therapy, causing elevated peripheral epinephrine levels, are required for successful extinction learning [235,258–260]. Under stressful conditions, the vagus nerve responds to those elevations in epinephrine and signals the brain to rapidly store newly acquired memories while, as part of the parasympathetic nervous system, it inhibits the sympathetic responses [187]. This probably happens via vagal projections to the LC via the NTS. Once activated, the LC provides norepinephrine throughout the brain, including extinction learning relevant brain regions like the basolateral amygdala and prefrontal cortex. It has already been shown that VNS can increase norepinephrine levels in the amygdala as well as in the medial prefrontal cortex [261,262]. Interestingly, a study from 2013 identified a small subpopulation of noradrenergic neurons in the subcoeruleus that specifically targets the insular cortex [263], indicating that VNS might have a direct influence on pIC activity via noradrenergic signaling. Accordingly, in the final set of experiments, I could show an increase in pIC activity during VNS (**Figure 26**). It is a promising avenue for follow-up studies to determine whether this effect is really mediated by noradrenaline release. Importantly, I could observe VNS evoked activity increase only in a subgroup of animals (‘VNS responders), and these animals were showing a significantly faster extinction learning when VNS was applied during extinction in comparison to the ‘non-responders’. Note that it is unclear whether the VNS responders and non-responders correspond to low and high fear state animals, respectively. It is also uncertain whether in the non-responder group, the VNS cuff was just placed incorrectly, or if the absence of VNS evoked pIC activity was the actual response profile of the animal. Important to mention here is that the group sizes were very small (N=5 responders, N=8 non-responders) and thus, it remains subject of future experiments to repeat the experiment with increased group sizes and to re-analyze potential differences in VNS evoked pIC activity between low and high fear state animals and their effects of extinction learning.

Nevertheless, this set of experiments provides, for the first time, insights into how VNS can alter pIC activity and thereby influence extinction learning. Further, this study also provides relevant information for translational research: VNS has been FDA approved for the treatment of epilepsy for over two decades

now, and since 2005, it is also approved for treatment-resistant depression. Thus, there is growing interest in extending the mood improving effects of VNS on patients suffering from other psychiatric conditions. A pilot study from 2008 that analyzed the effects of VNS on treatment-resistant anxiety disorders has already found some evidence for acute and long-term improvement in some of the tested patients [264]. On the other hand, a more recent study analyzing the safety of VNS in depressed patients already mention the development of anxiety symptoms as one of the side effects [265]. Accordingly, my results raise additional caution to whether VNS is a safe treatment for anxiety patients as I could show that VNS might also have unwanted negative effects on extinction learning.

4.3 Working model of the interoceptive insula's role in fear extinction

In conclusion, the main findings of my thesis were the following:

1) The pIC state-dependently processes learned cues and aversive stimuli.
2) The pIC processes sustained anxiety states.
3) Optogenetic pIC silencing has bidirectional effects on extinction learning depending on the internal fear state.
4) The internal state might be transmitted to the insular cortex via the vagus nerve as VNS modulates pIC activity.
5) VNS has similar bidirectional and state-dependent effects on extinction learning.

Even though the results are seemingly paradoxical, they could be interpreted by referring to the Yerkes-Dodson law, [238] that was developed in 1908 to describe the relationship between stimulus strength and the rapidity of habit formation for "difficult" discrimination learning tasks in mice. It suggests that in difficult tasks, performance is best when an individual is under optimal arousal levels, and performance would be impaired under conditions above or below optimal arousal levels. Nowadays, this "inverted-U-shape" function is used in psychology and translational research, and it is generally applied to explain the relationship between arousal intensity and performance, with performance increasing with arousal to an optimal point, above or below which performance decreases.

Results from this study also hint towards an inverted-U-shape relationship, if we assume that the internal fear state represents the arousal level of an animal. Under control conditions (**Figure 27 left)**, animals express a spectrum of internal fear states (x-axis), but despite those differences in arousal, all animals show an equal extinction performance (y-axis). Optogenetic pIC inhibition seems to cause a shift of this U-shape to the right (**Figure 27 right**): animals that were in a previously too low fear state are now pushed

toward a higher arousal state into an optimal learning range, showing an extinction facilitation; accordingly, animals that were already in an elevated fear state are now pushed towards a too high arousal state, thus causing an impairment in extinction performance. Interestingly, the exact same explanation can be applied to the experiments where the vagus nerve was stimulated. Hence, there must be a common underlying mechanism between insular and vagal influence on extinction learning. Paradoxically, I found that VNS alone increases pIC activity (**Figure 26**) and not (as one would expect from optogenetic silencing experiments) a decrease. However, these experiments were conducted in a neutral, not fear-evoking context. Thus, it is possible that VNS effects on pIC activity are also state-dependent, and might have a different influence on pIC activity when the animal is in an aroused state (e.g. during fear extinction), but this remains to be determined in follow-up experiments.

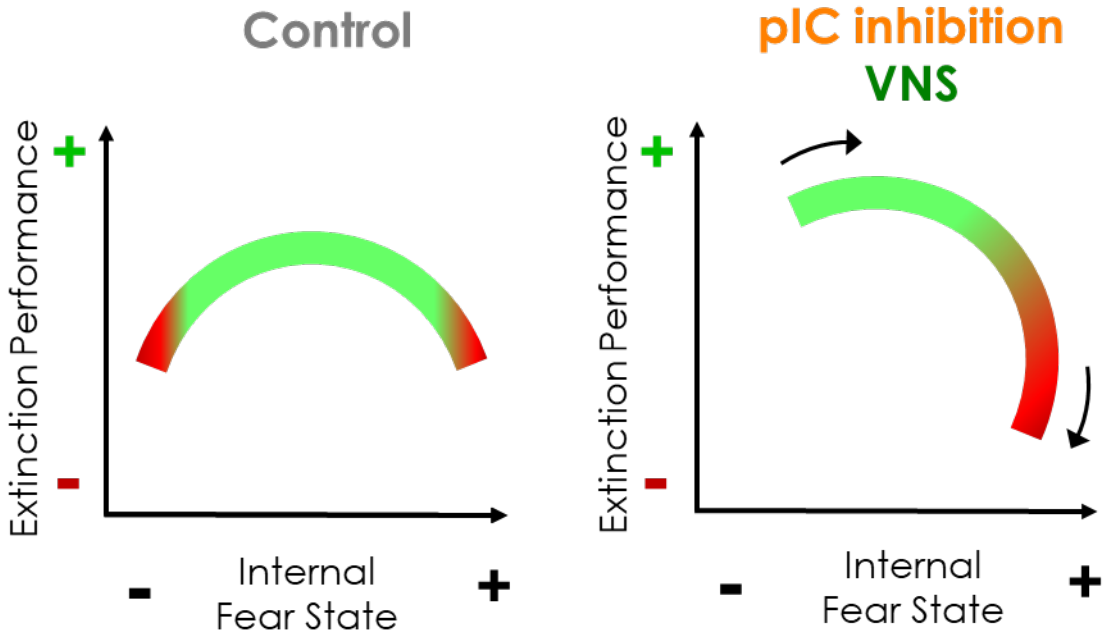


Figure 27: Working model of the interoceptive insula's role in fear extinction.
Left: Under physiological circumstances, the relationship between arousal intensity and performance can be described as an "inverted U-shape", with performance increasing with arousal to an optimal point, above or below which performance decreases. Right: Optogenetic pIC inhibition and VNS both seemingly cause a shift of this U-shape to the right: Animals that were in a previously too low fear state are now pushed toward a higher arousal state into an optimal learning range, showing an extinction facilitation; accordingly, animals that were already in an elevated fear state are now pushed towards a too high arousal state, thus causing an impairment in extinction performance.

As a result, it remains intriguing how pIC inhibition and VNS can have the same effects on extinction learning. However, a possible explanation how both manipulations can increase arousal level could be the following: as mentioned earlier, the IC is a site where bodily sensations and predictions about future bodily states converge. Thus, pIC inhibition during cues that predict an aversive outcome might cause a mismatch between interoceptive and prediction processes, thus eventually triggering an increase in arousal levels. Correspondingly, also VNS evoked manipulations of interoceptive signals during those cues might cause such a mismatch that elicits elevated arousal levels. Resulting alterations in arousal levels then in turn may influence extinction learning performance.

In conclusion, I suggest that the role of the pIC in extinction learning is a homeostatic maintenance of optimal performance levels despite differences in internal fear states, a process that is modulated by vagal afferents carrying interoceptive signals to the brain.

4.4 The pIC is playing a pivotal role in states of heightened anxiety

Despite the above-described findings on the interoceptive pIC's influence on extinction learning, I also extended previous evidence that the pIC is playing a more general role in states of heightened anxiety.

First, in a recent study from our lab, we have shown that the pIC does not only respond to acute aversive stimuli, but also encodes longer lasting aversive states [35]. As these data were acquired with a different method (two-photon imaging) using a different behavioral paradigm, I aimed to reproduce and extend those observations using single unit electrophysiology. Indeed, I could also identify subpopulation of neurons that encoded longer lasting fear states **(Figure 18**). These findings on emotional state-coding reveal interesting parallels with recent findings from the Andermann group [111] which showed that the insula reflects long lasting states of physiological need (hunger or thirst). Together, my own data, published data from our own group and others start to converge on the notion that the insula is important for the detection of aversive emotional and bodily states, to then in turn exert top-down control on motivated behaviors.

Further, a very interesting observation in this study was that fear-related elevated pIC activity does not simply reflect freezing behavior since FP data showed that during freezing, pIC activity was decreasing, and also in the single unit data, there was a large fraction of units that decreased their firing upon freezing onset. As we and other have shown that high fear states are correlated with high insular activity [35,87,88], one possible explanation for freezing evoked decrease in pIC activity is that freezing might be more an active instead of a passive coping strategy. Indeed, recent studies in humans and rodents have indicated

that freezing is not a passive state but rather a parasympathetic 'brake' on the otherwise active motor system, thereby enhancing perception and appropriate action preparation [266]. It is suggested that this freezing evoked parasympathetic 'brake' is driven by a connection between central amygdala (CeA) and ventro-lateral periaquaeductal gray (vlPAG), and our lab has recently demonstrated a strong unidirectional projection from the pIC to the CeA[35,41], which could be involved in the generation of this parasympathetic brake. Interestingly, in the FP data, I found that pIC activity seems to increase just before (~ 2 s) freezing onset ('pre-freezing' activity) (**Figure 11**), which is supported by observations from the single unit data that show similar changes in firing rate just before freezing onset (**Figure 17**). On a theoretical note, it is possible that when the fear state is increasing, also pIC activity increases, up until a certain "threshold" is reached that initiates freezing by pIC-CeA-vlPAG driven parasympathetic mechanisms. Freezing in turn then actively reduces insular activity, lowering the fear state and thus making way for improved perception and appropriate action preparation. To test this theory, it would be interesting to study if a "predictive signal" exists in the pIC that is able to predict fear behavior expression, but this remains subject to future studies.

Lastly, fiber photometry and electrophysiology pIC recordings in mice undergoing FC/Ext revealed a gradual in- and decrease of fear cue (CS+) evoked bulk pIC activity, or a gradual in- and decrease in the population size of fear cue responsive neurons, respectively (**Figure 8a and b, Figure 14**). Collectively, this suggests a fear-state dependent processing of salient cues. However, I could also always observe responses to the putatively neutral cue (CS-), for example in the FP experiments, there was a small, but apparent response to the putatively neutral cue (CS-) before any conditioning (during habituation) (**Figure 8c and d, Figure 13**). Additionally, also the single unit data showed an strong CS- response already during habituation, that was higher in amplitude and fraction in comparison to the CS+ response (which was pure tone) (**Figure 15**). This could on the one hand be explained by the physical properties of the CS- (constant white noise), as the pIC region where I was recording from is located closely to the insular auditory field (IAF) [48,267,268]. As the IAF is tonotopically organized [267], white noise might activate a large portion of this area. On the other hand, another possibility is that the CS- related white noise actually elicits an innate fear response in animals, even prior to any conditioning. Indeed, some studies have shown that white noise elicits greater physiological and behavioral responses than pure tones [269–271]. This is even reflected in my behavioral data, where animals usually showed a higher amount of freezing to the CS- in comparison to the CS+ during habituation (**Figure 6c and Figure 13**). Interestingly, in the fiber photometry experiments, CS- was partially even able to decrease pIC activity (**Supplementary figure 2**), and that was also reflected in decreasing freezing levels to the CS- during recall (**Supplementary figure 3**). This could be a hint towards the CS- acting as a so called 'safety signal' that is able to reduce fear even

below baseline levels [105,222,223], and more importantly, this also corroborated the hypothesis that CS evoked pIC activity reflects the fear state, rather than just pure sensory processing.

4.5 Limitations and outlook

4.5.1 Fiber photometry and electrophysiological single-unit recordings

First, regarding the electrophysiological single-unit recordings, one obvious limitation is that analysis was not explicitly differentiating between neuron subtypes (i.e. pyramidal cells and interneurons), which could have a major influence on the interpretation of the data. However, when I applied well-established methods to identify cell classes using extracellular waveform analysis [272], I could not readily discriminate two separate neuronal populations (**Supplementary figure 12**). This could be due to suboptimal recording or spike sorting approaches or, even more likely, because of a relatively low number of recorded neurons, resulting in only very few recorded interneurons. Thus, it will be crucial in follow-up experiments to either increase the group sizes, improve recording and unit identification techniques, or apply "Optotagging" methods (combining optogenetic neuronal activation with single-unit reocrdings[273]) to precisely and reliably identify units before/after recording sessions.

Further, it is important to note that I observed differenced in CS and freezing evoked responses between the two different recording techniques I employed, i.e. a large fraction of freezing activated units in electrophysiology recordings vs. a decrease of bulk activity during freezing in FP recording (**Figure 11 and Figure 17**). These differences might be explained by the characteristics of these distinct techniques. FP recordings only represent the net activity of the neuronal population, thus masking the possible differential activity in- or decrease of specific subpopulations. Furthermore, using a CaMKII promotor for GCaMP6s expression, my FP data exclusively derive from excitatory cells, while in single unit electrophysiology recordings I have included the activity of both excitatory and inhibitory cells, although I did not explicitly differentiate between those two main subpopulations in this study due to limited cell numbers (as mentioned above). Thus, subsequent studies are necessary to differentiate the contribution of excitatory and inhibitory cells to the observed state-dependent modulation of cue and freezing evoked pIC activity.

In 2018, a paper from the Torrealba group was published where they recorded single unit activity from rat pIC during fear extinction[97]. They identified two discrete groups of pIC neurons that were either what they termed 'freezing units' that were responsive to the CS+ during early extinction phases and become

silent during late extinction , or ‘extinction units’ that only became responsive at late extinction phases. Further, they also showed that the activity of both groups reflected CS+ evoked freezing behavior, which is in line with my findings. My data also confirm their findings of ‘freezing units’ as they probably correspond to the populations of CS+ responsive neurons that I have identified in early and late extinction, where the CS+ responsive population decreases from early to late extinction. However, I did not observe neurons corresponding to their ‘extinction units’, but this might be due to the way I analyzed the data as I did not compare the responses of the same neuron between early and late extinction. However, it is very likely that there are certain neurons that are ‘extinction’ or also ‘safety signal’ units, that become active once the fear state is decreasing, but this is subject for future analyses and experiment.

4.5.2 Optrode recordings

Optogenetic neuronal manipulation are nowadays a widely used tool. However, to validate this technique in our own laboratory was a crucial first step before starting the application in behavior experiment. Thus, I aimed to verify the ability of eNpHR3.0 to effectively inhibit neuronal activity of excitatory cells in the pIC (**Figure 19**). I found that a majority of recorded neurons (71%) were strongly, reliable and constantly reducing their firing rate during light application, confirming the effect of optogenetic neuronal silencing. Nevertheless, a small neuronal population (29%) was slightly activated during light application. However, this increase in firing rate was slower, weaker and less stable in comparison to the inhibited population. This observation was hinting towards an indirect activation of these neurons, maybe via a disinhibition mechanism (e.g. an excitatory neuron that is expressing eNpHR3.0 and projecting to an interneuron is inhibited, and thus this interneuron indirectly also decreases firing, then in turn causing an activation of another neuron). Nevertheless, the overall effect of optogenetic manipulation was a net decrease in activity, and therefore I decided to go forward with using this tool for subsequent behavioral experiments. Additionally, it has been discussed that the release of optogenetic silencing may cause rebound excitation[158]. In my experiments I did not explicitly observe the occurrence of such events, but this will be an important consideration for future studies. For example, there is a wealth of optogenetic silencing tools that use different mechanisms to hyperpolarize a neurons, for example Archaerhodopsin and its variants (proton pump) or iC++ and iChloC (anion-conducting Channelrhodopsins) [274], and one should consider the appropriate opsin for future experiments.

4.5.3 Behavioral experiments

An integral part of a good practice for differential fear conditioning paradigms is to counterbalance the auditory stimuli used for CS+ and CS- to control for the specificity of learned fear responses [270,275]. This is especially important in neuronal activity recording experiments, where auditory cue evoked neuronal responses are compared to differentiate between purely sensory evoked and acquired responses. However, I did not counterbalance the tones between CS+ and CS- in this study. I nevertheless could reliably show that even if animals showed an initial neuronal as well as behavioral response to CS- presentations, this response was gradually altered with learning. Nevertheless, counterbalancing the tones will be an important inclusion in future experiments.

Lastly, in recent years, it became more and more obvious that freezing is not the optimal readout for fear-related behavior. Automated freezing detection methods (as used in this study) are fast, but may incorrectly detect immobility or quiet resting as freezing behavior, whereas manual freezing detection methods are more precise, but very laborious. Additionally, mice may show multiple different fear responses, depending on the shock intensity[240], the cue intensity[276], and imminence of a threatening situation[277,278]. Those fear responses can, despite freezing, can also include risk assessment, defensive burying, ultrasonic vocalizations, fear potentiated startle, but also vegetative responses (i.e. relating to changes in autonomous nervous system such as heart rate, blood pressure of breathing frequency) [270]. Additionally, a recent study from our lab has shown that mice exhibit stereotyped facial expression that reflect the internal emotional state [279]. Thus, to get a more dynamic, comprehensive and precise readout of the current fear state of an animal, it will be worthwhile to combine several of those fear readout approaches in future experiments (for instance, recording facial expressions with a head-mounted camera and monitoring heart and breathing rate using a pulse oximeter).

5 APPENDICES

5.1 List of figures

5.2 References

1. Church, H. A., Corvin, A. & Lucey, J. V. Anxiety disorders. *Core Psychiatry* 207–224 (2012) doi:10.1016/B978-0-7020-3397-1.00016-1.
2. Kessler, R. C. *et al.* Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry* **6**, 168–76 (2007).
3. Stein, D. J. *et al.* Pharmacotherapy for post traumatic stress disorder (PTSD). *Cochrane database Syst. Rev.* **2006**, CD002795 (2006).
4. Penava, S. J., Otto, M. W., Pollack, M. H. & Rosenbaum, J. F. Current status of pharmacotherapy for PTSD: An effect size analysis of controlled studies. *Depress. Anxiety* **4**, 240–242 (1996).
5. Bandelow, B. *et al.* Efficacy of treatments for anxiety disorders. *Int. Clin. Psychopharmacol.* **30**, 183–192 (2015).
6. Foa, E. B., Keane, T. M., Friedman, M. J. & Cohen, J. A. *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies, 2nd ed. - PsycNET*. *American Psychological Association* (2009).
7. National Collaborating Centre for Mental Health. & National Institute for Clinical Excellence (Great Britain). *Post-traumatic stress disorder : the management of PTSD in adults and children in primary and secondary care*. (National Institute for Clinical Excellence, 2005).
8. Kaplan, J. S. & Tolin, D. F. Exposure Therapy for Anxiety Disorders. (2011).
9. Powers, M. B., Halpern, J. M., Ferenschak, M. P., Gillihan, S. J. & Foa, E. B. A meta-analytic review of prolonged exposure for posttraumatic stress disorder. *Clin. Psychol. Rev.* **30**, 635–641 (2010).
10. Hofmann, S. G. Enhancing exposure-based therapy from a translational research perspective. *Behav. Res. Ther.* **45**, 1987–2001 (2007).
11. Radulovic, J., Lee, R. & Ortony, A. State-Dependent Memory: Neurobiological Advances and Prospects for Translation to Dissociative Amnesia. *Front. Behav. Neurosci.* **12**, 259 (2018).
12. Pavlov, P. I. Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Oxford Univ. Press London, 1927.* (1927) doi:10.5214/ans.0972-7531.1017309.
13. Myers, K. M. & Davis, M. Mechanisms of fear extinction. *Mol. Psychiatry* **12**, 120–150 (2007).
14. McAllister, W. R., McAllister, D. E., Scoles, M. T. & Hampton, S. R. Persistence of fear-reducing behavior: relevance for the conditioning theory of neurosis. *J. Abnorm. Psychol.* **95**, 365–72 (1986).
15. Myers, K. M. & Davis, M. Behavioral and Neural Analysis of Extinction. *Neuron* **36**, 567–584

(2002).

16. Vervliet, B., Craske, M. G. & Hermans, D. Fear Extinction and Relapse: State of the Art. *Annu. Rev. Clin. Psychol.* **9**, 215–248 (2013).
17. Quirk, G. J. & Mueller, D. Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* **33**, 56–72 (2008).
18. Maren, S. & Holmes, A. Stress and Fear Extinction. *Neuropsychopharmacology* **41**, 58–79 (2016).
19. Kim, J. J. & Jung, M. W. Neural circuits and mechanisms involved in Pavlovian fear conditioning: a critical review. *Neurosci. Biobehav. Rev.* **30**, 188–202 (2006).
20. Maren, S. Neurobiology of Pavlovian Fear Conditioning. *Annu. Rev. Neurosci.* **24**, 897–931 (2001).
21. Kurth, F., Zilles, K., Fox, P. T., Laird, A. R. & Eickhoff, S. B. A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Struct. Funct.* **214**, 519–534 (2010).
22. Downar, J., Blumberger, D. M. & Daskalakis, Z. J. The Neural Crossroads of Psychiatric Illness: An Emerging Target for Brain Stimulation. *Trends Cogn. Sci.* **20**, 107–120 (2016).
23. Simmons, W. K. *et al.* Keeping the body in mind: insula functional organization and functional connectivity integrate interoceptive, exteroceptive, and emotional awareness. *Hum. Brain Mapp.* **34**, 2944–58 (2013).
24. Singer, T., Critchley, H. D. & Preuschoff, K. A common role of insula in feelings, empathy and uncertainty. *Trends Cogn. Sci.* **13**, 334–340 (2009).
25. Paulus, M. P. & Stein, M. B. An insular view of anxiety. *Biol. Psychiatry* **60**, 383–7 (2006).
26. Craig, A. D. B. How do you feel--now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* **10**, 59–70 (2009).
27. Craig, A. D. How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* **3**, 655–666 (2002).
28. Yiannakas, A. & Rosenblum, K. The Insula and Taste Learning. *Front. Mol. Neurosci.* **10**, 335 (2017).
29. Yamamoto, T., Yuyama, N., Kato, T. & Kawamura, Y. Gustatory responses of cortical neurons in rats. III. Neural and behavioral measures compared. *J. Neurophysiol.* **53**, 1370–86 (1985).
30. Yamamoto, T., Matsuo, R., Kiyomitsu, Y. & Kitamura, R. Taste responses of cortical neurons in freely ingesting rats. *J. Neurophysiol.* **61**, 1244–58 (1989).
31. Kosar, E. *et al.* Gustatory cortex in the rat. I. Physiological properties and cytoarchitecture. *Brain Res.* **379**, 329–41 (1986).
32. Cechetto, D. F. D. F. & Saper, C. C. B. Evidence for a viscerotopic sensory representation in the

cortex and thalamus in the rat. *J. Comp. Neurol.* **262**, 27–45 (1987).

33. Ogawa, H., Ito, S., Murayama, N. & Hasegawa, K. Taste area in granular and dysgranular insular cortices in the rat identified by stimulation of the entire oral cavity. *Neurosci. Res.* **9**, 196–201 (1990).
34. Hanamori, T., Kunitake, T., Kato, K. & Kannan, H. Responses of Neurons in the Insular Cortex to Gustatory, Visceral, and Nociceptive Stimuli in Rats. *J Neurophysiol* **79**, 2535–2545 (1998).
35. Gehrlach, D. A. *et al.* Aversive state processing in the posterior insular cortex. *Nat. Neurosci.* **22**, 1424–1437 (2019).
36. Butti, C. & Hof, P. R. The insular cortex: a comparative perspective. *Brain Struct. Funct.* **214**, 477–93 (2010).
37. Shi, C. J. & Cassell, M. D. Cortical, thalamic, and amygdaloid connections of the anterior and posterior insular cortices. *J. Comp. Neurol.* **399**, 440–68 (1998).
38. Mesulam, M.-M. & Mufson, E. J. Insula of the old world monkey. Architectonics in the insulo-orbito-temporal component of the paralimbic brain. *J. Comp. Neurol.* **212**, 1–22 (1982).
39. Cechetto, D. F. & Saper, C. B. Evidence for a viscerotopic sensory representation in the cortex and thalamus in the rat. *J. Comp. Neurol.* **262**, 27–45 (1987).
40. Allen, G. V., Saper, C. B., Hurley, K. M. & Cechetto, D. F. Organization of visceral and limbic connections in the insular cortex of the rat. *J. Comp. Neurol.* **311**, 1–16 (1991).
41. Gehrlach, D. A. *et al.* A whole-brain connectivity map of mouse insular cortex. *bioRxiv* 2020.02.10.941518 (2020) doi:10.1101/2020.02.10.941518.
42. Gogolla, N. The insular cortex. *Curr. Biol.* **27**, R580–R586 (2017).
43. Cechetto, D. F., JUN CHEN John, S. P., JUN CHEN Subcortical, S., Jun Chen, S. & Chen, S. J. Subcortical sites mediating sympathetic responses from insular cortex in rats. *Am. J. Physiol. Integr. Comp. Physiol.* **258**, R245–R255 (1990).
44. van der Kooy, D., Koda, L. Y., McGinty, J. F., Gerfen, C. R. & Bloom, F. E. The organization of projections from the cortes, amygdala, and hypothalamus to the nucleus of the solitary tract in rat. *J. Comp. Neurol.* **224**, 1–24 (1984).
45. Shipley, M. T. Insular cortex projection to the nucleus of the solitary tract and brainstem visceromotor regions in the mouse. *Brain Res. Bull.* **8**, 139–48 (1982).
46. Saper, C. B. Convergence of autonomic and limbic connections in the insular cortex of the rat. *J. Comp. Neurol.* **210**, 163–73 (1982).
47. Saper, C. B. Reciprocal parabrachial-cortical connections in the rat. *Brain Res.* (1982) doi:10.1016/0006-8993(82)90493-0.
48. Gogolla, N., Takesian, A. E., Feng, G., Fagiolini, M. & Hensch, T. K. Sensory Integration in

Mouse Insular Cortex Reflects GABA Circuit Maturation. *Neuron* **83**, 894–905 (2014).

49. Rodgers, K. M., Benison, A. M., Klein, A. & Barth, D. S. Auditory, somatosensory, and multisensory insular cortex in the rat. *Cereb. Cortex* **18**, 2941–51 (2008).
50. Chen, X., Gabitto, M., Peng, Y., Ryba, N. J. P. & Zuker, C. S. A gustotopic map of taste qualities in the mammalian brain. *Science* **333**, 1262–6 (2011).
51. Reynolds, S. M. & Zahm, D. S. Specificity in the projections of prefrontal and insular cortex to ventral striatopallidum and the extended amygdala. *J. Neurosci.* **25**, 11757–67 (2005).
52. Ghaziri, J. *et al.* Subcortical structural connectivity of insular subregions. *Sci. Rep.* **8**, 8596 (2018).
53. Reynolds, S. M. & Zahm, D. S. Specificity in the Projections of Prefrontal and Insular Cortex to Ventral Striatopallidum and the Extended Amygdala. *J. Neurosci.* **25**, 11757–11767 (2005).
54. Namburi, P., Al-Hasani, R., Calhoon, G. G., Bruchas, M. R. & Tye, K. M. Architectural Representation of Valence in the Limbic System. *Neuropsychopharmacol. Off. Publ. Am. Coll. NeuropsNamburi, P., Al-Hasani, R., Calhoon, G. G., Bruchas, M. R., Tye, K. M. (2015). Archit. Represent. Val. Limbic Syst. Neuropsychopharmacol. Off.* (2015) doi:10.1038/npp.2015.358.
55. Floresco, S. B. The Nucleus Accumbens: An Interface Between Cognition, Emotion, and Action. *Annu. Rev. Psychol.* **66**, 25–52 (2015).
56. Jaramillo, A. A. *et al.* Functional role for suppression of the insular-striatal circuit in modulating interoceptive effects of alcohol. *Addict. Biol.* **23**, 1020–1031 (2018).
57. Naqvi, N. H. & Bechara, A. The insula and drug addiction: an interoceptive view of pleasure, urges, and decision-making. *Brain Struct. Funct.* **214**, 435–450 (2010).
58. Flynn, F. G. Anatomy of the insula functional and clinical correlates. *Aphasiology* **13**, 55–78 (1999).
59. Mesulam, M.-M. & Mufson, E. J. Insula of the old world monkey. III: Efferent cortical output and comments on function. *J. Comp. Neurol.* **212**, 38–52 (1982).
60. Fujita, S. *et al.* Cytoarchitecture-Dependent Decrease in Propagation Velocity of Cortical Spreading Depression in the Rat Insular Cortex Revealed by Optical Imaging. *Cereb. Cortex* **26**, 1580–9 (2016).
61. Ruggiero, D. A., Mraovitch, S., Granata, A. R., Anwar, M. & Reis, D. J. A role of insular cortex in cardiovascular function. *J. Comp. Neurol.* **257**, 189–207 (1987).
62. Yasui, Y., Breder, C. D., Safer, C. B. & Cechetto, D. F. Autonomic responses and efferent pathways from the insular cortex in the rat. *J. Comp. Neurol.* **303**, 355–374 (1991).
63. Khalsa, S. S. *et al.* Interoception and Mental Health: A Roadmap. *Biol Psychiatry Cogn Neurosci Neuroimaging* **3**, 501–513 (2018).
64. Nattie, E. & Li, A. Respiration and autonomic regulation and orexin. *Prog. Brain Res.* **198**, 25–46

(2012).

65. Critchley, H. D. D. & Harrison, N. A. A. Visceral Influences on Brain and Behavior. *Neuron* **77**, 624–638 (2013).
66. Fealey, R. D. *Interoception and autonomic nervous system reflexes thermoregulation*. *Handbook of Clinical Neurology* vol. 117 (Elsevier B.V., 2013).
67. Von Leupoldt, A., Chan, P.-Y. Y. S., Esser, R. W. & Davenport, P. W. Emotions and neural processing of respiratory sensations investigated with respiratory-related evoked potentials. *Psychosom. Med.* **75**, 244–252 (2013).
68. Simons, L. E., Elman, I. & Borsook, D. Psychological Processing in Chronic Pain: A Neural Systems Approach. *Neurosci Biobehav Rev* **0**, 61–78 (2014).
69. Stevenson, R. J., Mahmut, M. & Rooney, K. Individual differences in the interoceptive states of hunger, fullness and thirst. *Appetite* (2015) doi:10.1016/j.appet.2015.06.008.
70. Oppenheimer, S. & Cechetto, D. The Insular Cortex and the Regulation of Cardiac Function. in *Comprehensive Physiology* vol. 6 1081–1133 (John Wiley & Sons, Inc., 2016).
71. Saper, C. B. The Central Autonomic Nervous System: Conscious Visceral Perception and Autonomic Pattern Generation. *Annu. Rev. Neurosci.* **25**, 433–469 (2002).
72. Nieuwenhuys, R. The insular cortex: a review. *Prog. Brain Res.* **195**, 123–63 (2012).
73. Dunitz, M. Vagus nerve stimulation. in *Introduction to Epilepsy* 441 (Taylor & Francis Group, 2012). doi:10.1017/CBO9781139103992.087.
74. Ruffoli, R. *et al.* The chemical neuroanatomy of vagus nerve stimulation. *J. Chem. Neuroanat.* **42**, 288–296 (2011).
75. Câmara, R. & Griessenauer, C. J. Anatomy of the Vagus Nerve. *Nerves Nerve Inj.* 385–397 (2015) doi:10.1016/B978-0-12-410390-0.00028-7.
76. Aleksandrov, V. G. *et al.* The Role of the Insular Cortex in the Control of Visceral Functions. *Hum. Physiol.* **41**, 553–561 (2015).
77. Aleksandrov, V. G., Ivanova, T. G. & Aleksandrova, N. P. Prefrontal control of respiration. *J. Physiol. Pharmacol.* **58**, 17–23 (2007).
78. Etkin, A., Büchel, C. & Gross, J. J. The neural bases of emotion regulation. *Nat. Rev. Neurosci.* **16**, 693–700 (2015).
79. Phan, K. L., Wager, T., Taylor, S. F. & Liberzon, I. Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. *Neuroimage* **16**, 331–348 (2002).
80. Nagai, M., Kishi, K. & Kato, S. Insular cortex and neuropsychiatric disorders: A review of recent literature. *Eur. Psychiatry* **22**, 387–394 (2007).
81. Tovote, P., Fadok, J. P. & Lüthi, A. Neuronal circuits for fear and anxiety. *Nat. Rev. Neurosci.* **16**,

317–331 (2015).

82. Simmons, A., Matthews, S. C., Stein, M. B. & Paulus, M. P. Anticipation of emotionally aversive visual stimuli activates right insula. *Neuroreport* **15**, 2261–5 (2004).
83. Schunck, T. *et al.* Functional magnetic resonance imaging characterization of CCK-4-induced panic attack and subsequent anticipatory anxiety. *Neuroimage* **31**, 1197–1208 (2006).
84. Cameron, O. G., Zubieta, J. K., Grunhaus, L. & Minoshima, S. Effects of yohimbine on cerebral blood flow, symptoms, and physiological functions in humans. *Psychosom. Med.* **62**, 549–559 (2000).
85. Servan-Schreiber, D., Perlstein, W. M., Cohen, J. D. & Mintun, M. Selective Pharmacological Activation of Limbic Structures in Human Volunteers. *J. Neuropsychiatry Clin. Neurosci.* **10**, 148–159 (1998).
86. Jeong, H. *et al.* Altered functional connectivity in the fear network of firefighters with repeated traumatic stress. *Br. J. Psychiatry* **214**, 347–353 (2019).
87. Stein, M. B., Simmons, A. N., Feinstein, J. S. & Paulus, M. P. Increased Amygdala and Insula Activation During Emotion Processing in Anxiety-Prone Subjects. *Am. J. Psychiatry* **164**, 318–327 (2007).
88. Fullana, M. A. *et al.* Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol. Psychiatry* **21**, 500–508 (2016).
89. Greco, J. A. & Liberzon, I. Neuroimaging of Fear-Associated Learning. *Neuropsychopharmacology* **41**, 320–334 (2016).
90. Rauch, S. L., Savage, C. R., Alpert, N. M., Fischman, A. J. & Jenike, M. A. The functional neuroanatomy of anxiety: A study of three disorders using positron emission tomography and symptom provocation. *Biol. Psychiatry* **42**, 446–452 (1997).
91. Etkin, A. & Wager, T. D. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatry* **164**, 1476–88 (2007).
92. Damsa, C., Kosel, M. & Moussally, J. Current status of brain imaging in anxiety disorders. *Curr. Opin. Psychiatry* **22**, 96–110 (2009).
93. Shin, L. M. & Liberzon, I. The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology* **35**, 169–91 (2010).
94. Goossens, L., Sunaert, S., Peeters, R., Griez, E. J. L. & Schruers, K. R. J. Amygdala Hyperfunction in Phobic Fear Normalizes After Exposure. *Biol. Psychiatry* **62**, 1119–1125 (2007).
95. Hoehn-Saric, R., Schlund, M. W. & Wong, S. H. . Effects of citalopram on worry and brain activation in patients with generalized anxiety disorder. *Psychiatry Res. Neuroimaging* **131**, 11–

21 (2004).

96. Lázaro, L. *et al.* Cerebral activation in children and adolescents with obsessive–compulsive disorder before and after treatment: A functional MRI study. *J. Psychiatr. Res.* **42**, 1051–1059 (2008).
97. Casanova, J. P., Aguilar-Rivera, M., Rodríguez, M. de los Á., Coleman, T. P. & Torrealba, F. The activity of discrete sets of neurons in the posterior insula correlates with the behavioral expression and extinction of conditioned fear. *J. Neurophysiol.* **120**, 1906–1913 (2018).
98. Rodríguez, M. *et al.* Interoceptive Insular Cortex Mediates Both Innate Fear and Contextual Threat Conditioning to Predator Odor. *Front. Behav. Neurosci.* **13**, 283 (2020).
99. Méndez-Ruette, M. *et al.* The Role of the Rodent Insula in Anxiety. *Front. Physiol.* **10**, 330 (2019).
100. Li, H., Chen, L., Li, P., Wang, X. & Zhai, H. Insular muscarinic signaling regulates anxiety-like behaviors in rats on the elevated plus-maze. *Behav. Brain Res.* **270**, 256–260 (2014).
101. Brunzell, D. H., Kim, J. J. & Brunzell, Darlene H.; Kim, J. J. Fear conditioning to tone, but not to context, is attenuated by lesions of the insular cortex and posterior extension of the intralaminar complex in rats. *Behav. Neurosci.* **115**, 365–75 (2001).
102. Alves, F. H. F. *et al.* Involvement of the insular cortex in the consolidation and expression of contextual fear conditioning. *Eur. J. Neurosci.* **38**, 2300–7 (2013).
103. Casanova, J. P. *et al.* A role for the interoceptive insular cortex in the consolidation of learned fear. *Behav. Brain Res.* **296**, 70–77 (2015).
104. Lanuza, E., Nader, K. & Ledoux, J. . Unconditioned stimulus pathways to the amygdala: effects of posterior thalamic and cortical lesions on fear conditioning. *Neuroscience* **125**, 305–315 (2004).
105. Foilb, A. R., Flyer-Adams, J. G., Maier, S. F. & Christianson, J. P. Posterior insular cortex is necessary for conditioned inhibition of fear. *Neurobiol. Learn. Mem.* (2016) doi:10.1016/j.nlm.2016.08.004.
106. Christianson, J. P. *et al.* Safety signals mitigate the consequences of uncontrollable stress via a circuit involving the sensory insular cortex and bed nucleus of the stria terminalis. *Biol. Psychiatry* **70**, 458–64 (2011).
107. Christianson, J. P. *et al.* The sensory insular cortex mediates the stress-buffering effects of safety signals but not behavioral control. *J. Neurosci.* **28**, 13703–11 (2008).
108. Kong, E., Monje, F. J., Hirsch, J. & Pollak, D. D. Learning not to fear: neural correlates of learned safety. *Neuropsychopharmacology* **39**, 515–27 (2014).
109. Berridge, K. C., Ho, C.-Y., Richard, J. M. & DiFeliceantonio, A. G. The tempted brain eats: pleasure and desire circuits in obesity and eating disorders. *Brain Res.* **1350**, 43–64 (2010).
110. Livneh, Y. *et al.* Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature*

(2017) doi:10.1038/nature22375.

111. Livneh, Y. *et al.* Estimation of Current and Future Physiological States in Insular Cortex. *Neuron* **105**, 1–18 (2020).
112. Meier, L. *et al.* Thirst-Dependent Activity of the Insular Cortex Reflects its Emotion-Related Subdivision: A Cerebral Blood Flow Study. *Neuroscience* **383**, 170–177 (2018).
113. Tataranni, P. A. *et al.* Neuroanatomical correlates of hunger and satiation in humans using positron emission tomography. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4569–74 (1999).
114. Egan, G. *et al.* Neural correlates of the emergence of consciousness of thirst. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15241–15246 (2003).
115. Paulus, M. P., Feinstein, J. S. & Khalsa, S. S. An Active Inference Approach to Interoceptive Psychopathology. *Annu. Rev. Clin. Psychol.* **15**, 97–122 (2019).
116. Barrett, L. F. & Simmons, W. K. Interoceptive predictions in the brain. *Nat. Rev. Neurosci.* **16**, 419–429 (2015).
117. Maren, S., Phan, K. L. & Liberzon, I. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nat Rev Neurosci.* **14**, 417–428 (2013).
118. Quadt, L., Critchley, H. D. & Garfinkel, S. N. The neurobiology of interoception in health and disease. *Ann. N. Y. Acad. Sci.* **1428**, 112–128 (2018).
119. Gardner, M. P. H. & Fontanini, A. Encoding and tracking of outcome-specific expectancy in the gustatory cortex of alert rats. *J. Neurosci.* **34**, 13000–17 (2014).
120. Kusumoto-Yoshida, I., Liu, H., Chen, B. T., Fontanini, A. & Bonci, A. Central role for the insular cortex in mediating conditioned responses to anticipatory cues. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6–11 (2015).
121. Paulus, M. P. & Stein, M. B. Interoception in anxiety and depression. *Brain Struct. Funct.* **214**, 1–13 (2010).
122. Roth, W. T. *et al.* Imipramine and alprazolam effects on stress test reactivity in panic disorder. *Biol. Psychiatry* **31**, 35–51 (1992).
123. Ehlers, A., Mayou, R. A., Sprigings, D. C. & Birkhead, J. Psychological and Perceptual Factors Associated With Arrhythmias and Benign Palpitations. *Psychosom. Med.* **62**, 693–702 (2000).
124. Pineles, S. L. & Mineka, S. Attentional Biases to Internal and External Sources of Potential Threat in Social Anxiety. *J. Abnorm. Psychol.* **114**, 314–318 (2005).
125. Van der Does, A. J. W., Antony, M. M., Ehlers, A. & Barsky, A. J. Heartbeat perception in panic disorder: a reanalysis. *Behav. Res. Ther.* **38**, 47–62 (2000).
126. Wald, J. & Taylor, S. Interoceptive Exposure Therapy Combined with Trauma-related Exposure Therapy for Post-traumatic Stress Disorder: a Case Report. *Cogn. Behav. Ther.* **34**, 34–40 (2005).

127. White, K. S., Brown, T. A., Somers, T. J. & Barlow, D. H. Avoidance behavior in panic disorder: The moderating influence of perceived control. *Behav. Res. Ther.* **44**, 147–157 (2006).
128. Zoellner, L. A. & Craske, M. G. Interoceptive accuracy and panic. *Behav. Res. Ther.* **37**, 1141–58 (1999).
129. Dalton, K. M., Kalin, N. H., Grist, T. M. & Davidson, R. J. Neural-Cardiac Coupling in Threat-Evoked Anxiety. *J. Cogn. Neurosci.* **17**, 969–980 (2005).
130. Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A. & Dolan, R. J. Neural systems supporting interoceptive awareness. *Nat. Neurosci.* **7**, 189–195 (2004).
131. Pollatos, O., Kirsch, W. & Schandry, R. On the relationship between interoceptive awareness, emotional experience, and brain processes. *Cogn. Brain Res.* **25**, 948–962 (2005).
132. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* vol. 93 480–490 (2017).
133. Logothetis, N. K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).
134. Bronstein, J. *et al.* Deep Brain Stimulation for Parkinson Disease: An Expert Consensus and Review of Key Issues. *Arch. Neurol.* **68**, (2011).
135. Teegarden, S. Behavioral Phenotyping in Rats and Mice. *Mater. Methods* **2**, (2012).
136. O'Neil, M. F. & Moore, N. A. Animal models of depression: are there any? *Hum. Psychopharmacol. Clin. Exp.* **18**, 239–254 (2003).
137. Barad, M. Fear extinction in rodents: basic insight to clinical promise. *Curr. Opin. Neurobiol.* **15**, 710–715 (2005).
138. Chang, C. *et al.* Fear extinction in rodents. *Curr. Protoc. Neurosci.* **Chapter 8**, Unit8.23 (2009).
139. Singewald, N. & Holmes, A. Rodent models of impaired fear extinction. *Psychopharmacology (Berl).* **236**, 21 (2019).
140. Walf, A. A. & Frye, C. A. The use of the elevated plus maze as an assay of anxiety-related behavior in rodents. *Nat. Protoc.* **2**, 322–8 (2007).
141. Walsh, R. N. & Cummins, R. A. The open-field test: A critical review. *Psychol. Bull.* **83**, 482–504 (1976).
142. Herry, C. & Johansen, J. P. Encoding of fear learning and memory in distributed neuronal circuits. *Nat. Neurosci.* **17**, 1644–1654 (2014).
143. Marek, R. & Sah, P. Neural Circuits Mediating Fear Learning and Extinction. in 35–48 (Springer, Cham, 2018). doi:10.1007/978-3-319-94593-4_2.
144. Milad, M. R. & Quirk, G. J. Fear Extinction as a Model for Translational Neuroscience: Ten Years of Progress. *Annu. Rev. Psychol.* **63**, 129–151 (2012).
145. Rauch, S. L., Shin, L. M. & Phelps, E. A. Neurocircuitry Models of Posttraumatic Stress Disorder

and Extinction: Human Neuroimaging Research—Past, Present, and Future. *Biol. Psychiatry* **60**, 376–382 (2006).

146. P, D. *et al.* Updated Meta-Analysis of Classical Fear Conditioning in the Anxiety Disorders. *Depress. Anxiety* **32**, (2015).
147. Wicking, M. *et al.* Deficient fear extinction memory in posttraumatic stress disorder. *Neurobiol. Learn. Mem.* **136**, 116–126 (2016).
148. Forcadell, E. *et al.* Does Fear Extinction in the Laboratory Predict Outcomes of Exposure Therapy? A Treatment Analog Study. *Int. J. Psychophysiol.* **121**, (2017).
149. Deisseroth, K. Optogenetics. *Nat. Methods* **8**, 26–29 (2011).
150. Fenno, L., Yizhar, O. & Deisseroth, K. The development and application of optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
151. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–8 (2005).
152. Mattis, J. *et al.* Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. *Nat. Methods* **9**, 159–72 (2012).
153. Gradinaru, V. *et al.* Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154–65 (2010).
154. Zhang, F. *et al.* Table 2 : Optogenetic interrogation of neural circuits: technology for probing mammalian brain structures : Nature Protocols. *Nat. Protoc.* **5**, 439–56 (2010).
155. Shirai, F. & Hayashi-Takagi, A. Optogenetics: Applications in psychiatric research. *Psychiatry Clin. Neurosci.* **71**, 363–372 (2017).
156. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in neural systems. *Neuron* **71**, 9–34 (2011).
157. Häusser, M. Optogenetics: the age of light. *Nat. Methods* **11**, 1012–1014 (2014).
158. Kravitz, A. V. & Bonci, A. Optogenetics, physiology, and emotions. *Front. Behav. Neurosci.* **7**, 169 (2013).
159. Tank, D., Sugimori, M., Connor, J. & Llinas, R. Spatially resolved calcium dynamics of mammalian Purkinje cells in cerebellar slice. *Science (80-.).* **242**, 773–777 (1988).
160. Baker, P. F., Hodgkin, A. L. & Ridgway, E. B. Depolarization and calcium entry in squid giant axons. *J. Physiol.* **218**, 709–755 (1971).
161. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
162. Helmchen, F. & Denk, W. Deep tissue two-photon microscopy. *Nature Methods* vol. 2 932–940 (2005).

163. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science (80-.).* **248**, 73–76 (1990).

164. Aharoni, D. & Hoogland, T. M. Circuit investigations with open-source miniaturized microscopes: Past, present and future. *Frontiers in Cellular Neuroscience* vol. 13 (2019).

165. Gunaydin, L. A. A. *et al.* Natural Neural Projection Dynamics Underlying Social Behavior. *Cell* **157**, 1535–1551 (2014).

166. Cui, G. *et al.* Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature* **494**, 238–242 (2013).

167. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).

168. Vandecasteele, M. *et al.* Large-scale Recording of Neurons by Movable Silicon Probes in Behaving Rodents. *J. Vis. Exp.* (2012) doi:10.3791/3568.

169. Buzsaki, G. *Rhythms of the Brain*. (2006).

170. Csicsvari, J. *et al.* Massively Parallel Recording of Unit and Local Field Potentials With Silicon-Based Electrodes. *J. Neurophysiol.* **90**, (2003).

171. G, B. Large-scale Recording of Neuronal Ensembles. *Nat. Neurosci.* **7**, (2004).

172. Obien, M. E. J., Deligkaris, K., Bullmann, T., Bakkum, D. J. & Frey, U. Revealing neuronal function through microelectrode array recordings. *Front. Neurosci.* **8**, 423 (2014).

173. Heinricher, M. Principles of Extracellular Single-Unit Recording. *Microelectrode Rec. Mov. Disord. Surg.* **c**, 8–13 (2014).

174. Anastassiou, C. A., Perin, R., Buzsáki, G., Markram, H. & Koch, C. Cell type- and activity-dependent extracellular correlates of intracellular spiking. *J. Neurophysiol.* **114**, 608–23 (2015).

175. Barthó, P. *et al.* Characterization of Neocortical Principal Cells and Interneurons by Network Interactions and Extracellular Features. *J. Neurophysiol.* **92**, (2004).

176. Trainito, C., von Nicolai, C., Miller, E. K. & Siegel, M. Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex. *Curr. Biol.* **29**, 2973-2982.e5 (2019).

177. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–41 (2016).

178. Lee, J. *et al.* YASS: Yet Another Spike Sorter. *bioRxiv* (2017).

179. Yger, P. *et al.* A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. *Elife* **7**, (2018).

180. Caro-Martín, C. R., Delgado-García, J. M., Gruart, A. & Sánchez-Campusano, R. Spike sorting based on shape, phase, and distribution features, and K-TOPS clustering with validity and error indices. *Sci. Rep.* **8**, 17796 (2018).

181. Bestel, R., Daus, A. W. & Thielemann, C. A novel automated spike sorting algorithm with

adaptable feature extraction. *J. Neurosci. Methods* **211**, 168–178 (2012).

182. Leibig, C., Wachtler, T. & Zeck, G. Unsupervised neural spike sorting for high-density microelectrode arrays with convolutive independent component analysis. *J. Neurosci. Methods* **271**, 1–13 (2016).
183. Fournier, J., Mueller, C. M., Shein-Idelson, M., Hemberger, M. & Laurent, G. Consensus-Based Sorting of Neuronal Spike Waveforms. *PLoS One* **11**, (2016).
184. Rey, H. G., Pedreira, C. & Quian Quiroga, R. Past, present and future of spike sorting techniques. *Brain Res. Bull.* **119**, 106–117 (2015).
185. Jahanmiri-Nezhad, F., Barkhaus, P. E., Rymer, W. Z. & Zhou, P. Spike sorting paradigm for classification of multi-channel recorded fasciculation potentials. *Comput. Biol. Med.* **55**, 26–35 (2014).
186. Noller, C. M., Levine, Y. A., Urakov, T. M., Aronson, J. P. & Nash, M. S. Vagus Nerve Stimulation in Rodent Models: An Overview of Technical Considerations. *Front. Neurosci.* **13**, (2019).
187. Noble, L. J., Souza, R. R. & McIntyre, C. K. Vagus nerve stimulation as a tool for enhancing extinction in exposure-based therapies. *Psychopharmacology* vol. 236 355–367 (2019).
188. Zagon, A. Does the vagus nerve mediate the sixth sense? *Trends Neurosci.* **24**, 671–673 (2001).
189. Howland, R. H. Vagus Nerve Stimulation. *Curr. Behav. Neurosci. reports* **1**, 64 (2014).
190. Berthoud, H. R. & Neuhuber, W. L. Functional and chemical anatomy of the afferent vagal system. *Auton. Neurosci.* **85**, 1–17 (2000).
191. Corning, J. L. Considerations on the patholoy and therapeutics of epilepsy. *J. Nerv. Ment. Dis.* **10**, 243–248 (1883).
192. Corning, J. L. Carotid Compresion and Brain Rest. in (ADF Randolph & Company, 1882).
193. Bailey, P. & Bremer, F. A SENSORY CORTICAL REPRESENTATION OF THE VAGUS NERVE: WITH A NOTE ON THE EFFECTS OF LOW BLOOD PRESSURE ON THE CORTICAL ELECTROGRAM. *J. Neurophysiol.* **1**, 405–412 (1938).
194. Dell, P. & Olson, R. Secondary Mesencephalic, Diencephalic and Amygdalian Projections of Vagal Visceral Afferences. *C R Seances Soc Biol Fil* 13–14 (1951).
195. Zabara, J. Peripheral control of hypersynchronous discharge in epilepsy. *Electroencephalogr. Clin. Neurophysiol.* **61**, S162 (1985).
196. Zabara, J. Time course of seizure control to breaf repetitive stimuli. *Epilepsia* **26**, 518 (1985).
197. Zabara, J. Inhibition of Experimental Seizures in Canines by Repetitive Vagal Stimulation. *Epilepsia* **33**, 1005–1012 (1992).
198. Penry, J. K. & Dean, J. C. Prevention of Intractable Partial Seizures by Intermittent Vagal

Stimulation in Humans: Preliminary Results. *Epilepsia* **31**, S40–S43 (1990).

199. Ben-Menachem, E. *et al.* Vagus Nerve Stimulation for Treatment of Partial Seizures: 1. A Controlled Study of Effect on Seizures. *Epilepsia* **35**, 616–626 (1994).
200. Handforth, A. *et al.* A randomized controlled trial of chronic vagus nerve stimulation for treatment of medically intractable seizures. The Vagus Nerve Stimulation Study Group. *Neurology* **45**, 224–30 (1995).
201. Henry, T. R. *et al.* Brain Blood Flow Alterations Induced by Therapeutic Vagus Nerve Stimulation in Partial Epilepsy: I. Acute Effects at High and Low Levels of Stimulation. *Epilepsia* **39**, 983–990 (1998).
202. Rush, A. J. *et al.* Vagus nerve stimulation (VNS) for treatment-resistant depressions: a multicenter study. *Biol. Psychiatry* **47**, 276–286 (2000).
203. Howland, R. H., Shutt, L. S., Berman, S. R., Spotts, C. R. & Denko, T. The emerging use of technology for the treatment of depression and other neuropsychiatric disorders. *Ann. Clin. Psychiatry* **23**, 48–62 (2011).
204. George, M. S. *et al.* A pilot study of vagus nerve stimulation (VNS) for treatment-resistant anxiety disorders. *Brain Stimul.* **1**, 112–21 (2008).
205. Shah, A. P., Carreno, F. R., Wu, H., Chung, Y. A. & Frazer, A. Role of TrkB in the anxiolytic-like and antidepressant-like effects of vagal nerve stimulation: Comparison with desipramine. *Neuroscience* **322**, 273–86 (2016).
206. Noble, L. J. *et al.* Vagus nerve stimulation promotes generalization of conditioned fear extinction and reduces anxiety in rats. *Brain Stimul.* **12**, 9–18 (2019).
207. Mathew, E. *et al.* Vagus nerve stimulation produces immediate dose-dependent anxiolytic effect in rats. *J. Affect. Disord.* **265**, 552–557 (2020).
208. Alvarez-Dieppa, A. C., Griffin, K., Cavalier, S. & McIntyre, C. K. Vagus Nerve Stimulation Enhances Extinction of Conditioned Fear in Rats and Modulates Arc Protein, CaMKII, and GluN2B-Containing NMDA Receptors in the Basolateral Amygdala. *Neural Plast.* **2016**, 4273280 (2016).
209. Souza, R. R. *et al.* Vagus nerve stimulation reverses the extinction impairments in a model of PTSD with prolonged and repeated trauma. *Stress* 1–12 (2019) doi:10.1080/10253890.2019.1602604.
210. Noble, L. J., Chuah, A., Callahan, K. K., Souza, R. R. & McIntyre, C. K. Peripheral effects of vagus nerve stimulation on anxiety and extinction of conditioned fear in rats. *Learn. Mem.* **26**, 245–251 (2019).
211. Noble, L. J. *et al.* Effects of vagus nerve stimulation on extinction of conditioned fear and post-

traumatic stress disorder symptoms in rats. *Transl. Psychiatry* **7**, e1217 (2017).

212. Peña, D. F., Engineer, N. D. & McIntyre, C. K. Rapid remission of conditioned fear expression with extinction training paired with vagus nerve stimulation. *Biol. Psychiatry* **73**, 1071–1077 (2013).
213. Peña, D. F. *et al.* Vagus nerve stimulation enhances extinction of conditioned fear and modulates plasticity in the pathway from the ventromedial prefrontal cortex to the Amygdala. *Front. Behav. Neurosci.* **8**, 327 (2014).
214. Klarer, M. *et al.* Gut vagal afferents differentially modulate innate anxiety and learned fear. *J. Neurosci.* **34**, 7067–76 (2014).
215. Akerman, S. & Romero-Reyes, M. Vagus Nerve Stimulation. in 87–98 (Springer, Cham, 2020). doi:10.1007/978-3-030-14121-9_6.
216. Lerner, T. N. *et al.* Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* **162**, 635–647 (2015).
217. Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates, 4th Edition | George Paxinos, Keith Franklin | ISBN 9780123910578. http://store.elsevier.com/Paxinos-and-Franklins-the-Mouse-Brain-in-Stereotaxic-Coordinates/George-Paxinos/isbn-9780123910578/.
218. Starr, C. J. *et al.* Roles of the insular cortex in the modulation of pain: insights from brain lesions. *J. Neurosci.* **29**, 2684–94 (2009).
219. Lu, C. *et al.* Insular Cortex is Critical for the Perception, Modulation, and Chronification of Pain. *Neurosci. Bull.* **32**, 191–201 (2016).
220. Dana, H. *et al.* High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657 (2019).
221. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
222. Pollak, D. D., Monje, F. J. & Lubec, G. The learned safety paradigm as a mouse model for neuropsychiatric research. *Nat. Protoc.* **5**, 954–962 (2010).
223. Sangha, S., Diehl, M. M., Bergstrom, H. C. & Drew, M. R. Know safety, no fear. *Neurosci. Biobehav. Rev.* (2019) doi:10.1016/j.neubiorev.2019.11.006.
224. Dejean, C. *et al.* Prefrontal neuronal assemblies temporally control fear behaviour. *Nature* (2016) doi:10.1038/nature18630.
225. Karalis, N. *et al.* 4-Hz oscillations synchronize prefrontal–amygdala circuits during fear behavior. *Nat. Neurosci.* **19**, 605–612 (2016).
226. dos Santos Corrêa, M. *et al.* Relationship between footshock intensity, post-training corticosterone release and contextual fear memory specificity over time. *Psychoneuroendocrinology* **110**, 104447

(2019).

227. Lindau, M., Almkvist, O. & Mohammed, A. H. *Effects of Stress on Learning and Memory*. *Stress: Concepts, Cognition, Emotion, and Behavior* 153–160 (2016). doi:10.1016/B978-0-12-800951-2.00018-2.
228. Salehi, B., Cordero, M. I. & Sandi, C. Learning under stress: the inverted-U-shape function revisited. *Learn. Mem.* **17**, 522–30 (2010).
229. Sandi, C. & Pinelo-Nava, M. T. Stress and Memory: Behavioral Effects and Neurobiological Mechanisms. *Neural Plast.* **2007**, 1–20 (2007).
230. Ribeiro, R. L. *et al.* The "Anxiety State" and Its Relation with Rat Models of Memory and Habituation. *Neurobiol. Learn. Mem.* **72**, 78–94 (1999).
231. Maroun, M. & Akirav, I. Arousal and stress effects on consolidation and reconsolidation of recognition memory. *Neuropsychopharmacology* **33**, 394–405 (2008).
232. Rau, V., DeCola, J. P. & Fanselow, M. S. Stress-induced enhancement of fear learning: An animal model of posttraumatic stress disorder. *Neurosci. Biobehav. Rev.* **29**, 1207–1223 (2005).
233. Raio, C. M. & Phelps, E. A. The influence of acute stress on the regulation of conditioned fear. *Neurobiol. Stress* **1**, 134–146 (2015).
234. Peyrot, C., Brouillard, A., Morand-Beaulieu, S. & Marin, M. A review on how stress modulates fear conditioning: Let's not forget the role of sex and sex hormones. *Behav. Res. Ther.* **129**, 103615 (2020).
235. Giustino, T. F., Maren, S., Sousa, N. & Norrholm, S. D. Noradrenergic Modulation of Fear Conditioning and Extinction. (2018) doi:10.3389/fnbeh.2018.00043.
236. Paul E. Gold. *Modulation of Emotional and Nonemotional Memories:Same Pharmacological Systems, Different Neuroanatomical Systems*. (1995).
237. Morley, K. C., Gallate, J. E., Hunt, G. E., Mallet, P. E. & McGregor, I. S. Increased anxiety and impaired memory in rats 3 months after administration of 3,4-methylenedioxymethamphetamine ('Ecstasy'). *Eur. J. Pharmacol.* **433**, 91–99 (2001).
238. Yerkes, R. M. & Dodson, J. D. The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* **18**, 459–482 (1908).
239. Shors, T. J. & Servatius, R. J. The Contribution of Stressor Intensity, Duration, and Context to the Stress-Induced Facilitation of Associative Learning. *Neurobiol. Learn. Mem.* **68**, 92–96 (1997).
240. Laxmi, T. R., Stork, O. & Pape, H.-C. Generalisation of conditioned fear and its behavioural expression in mice. *Behav. Brain Res.* **145**, 89–98 (2003).
241. Sandi, C. *et al.* Chronic stress-induced alterations in amygdala responsiveness and behavior - modulation by trait anxiety and corticotropin-releasing factor systems. *Eur. J. Neurosci.* **28**, 1836–

1848 (2008).

242. Luksys, G., Gerstner, W. & Sandi, C. Stress, genotype and norepinephrine in the prediction of mouse behavior using reinforcement learning. *Nat. Neurosci. 2009 129* **12**, 1180–1186 (2009).
243. Critchley, H. D. & Garfinkel, S. N. Interoception and emotion. *Curr. Opin. Psychol.* **17**, 7–14 (2017).
244. Spielberger, C. D. State-Trait Anxiety Inventory. in *The Corsini Encyclopedia of Psychology* 1–1 (John Wiley & Sons, Inc., 2010). doi:10.1002/9780470479216.corpsy0943.
245. Spielberger, C. D. *Understanding stress and anxiety.* (Harper & Row).
246. Spielberger, C. D. *Anxiety: Current trends in theory and research*. (1972).
247. Liebsch, G., Montkowski, A., Holsboer, F. & Landgraf, R. Behavioural profiles of two Wistar rat lines selectively bred for high or low anxiety-related behaviour. *Behav. Brain Res.* **94**, 301–310 (1998).
248. Muigg, P. *et al.* Impaired extinction of learned fear in rats selectively bred for high anxiety - evidence of altered neuronal processing in prefrontal-amygdala pathways. *Eur. J. Neurosci.* **28**, 2299–2309 (2008).
249. Sartori, S. B. & Singewald, N. New pharmacological strategies for augmenting extinction learning in anxiety disorders. (2017) doi:10.1515/nf-2017-A011.
250. Peña, D. F., Engineer VP of Preclinical Affairs, N. D., McIntyre, C. K., Professor of Cognition, A. & McIntyre, C. Rapid Remission of Conditioned Fear Expression with Extinction Training Paired with Vagus Nerve Stimulation. *Biol Psychiatry* **73**, 1071–1077 (2013).
251. Loerwald, K. W. *et al.* Varying stimulation parameters to improve cortical plasticity generated by VNS-tone pairing. *Neuroscience* **388**, 239 (2018).
252. Loerwald, K. W., Borland, M. S., Rennaker, R. L., Hays, S. A. & Kilgard, M. P. The interaction of pulse width and current intensity on the extent of cortical plasticity evoked by vagus nerve stimulation. *Brain Stimul.* **11**, 271–277 (2018).
253. Borland, M. S. *et al.* Cortical Map Plasticity as a Function of Vagus Nerve Stimulation Intensity. *Brain Stimul.* **9**, 117–123 (2016).
254. Clark, K. B., Krahl, S. E., Smith, D. C. & Jensen, R. A. Post-training unilateral vagal stimulation enhances retention performance in the rat. *Neurobiol. Learn. Mem.* **63**, 213–6 (1995).
255. Clark, K. B., Naritoku, D. K., Smith, D. C., Browning, R. A. & Jensen, R. A. Enhanced recognition memory following vagus nerve stimulation in human subjects. *Nat. Neurosci.* **2**, 94–8 (1999).
256. Clark, K. B. B. *et al.* Posttraining Electrical Stimulation of Vagal Afferents with Concomitant Vagal Efferent Inactivation Enhances Memory Storage Processes in the Rat. *Neurobiol. Learn. Mem.* **70**, 364–373 (1998).

257. Clark, K. B., Naritoku, D. K., Smith, D. C., Browning, R. A. & Jensen, R. A. Enhanced recognition memory following vagus nerve stimulation in human subjects. *Nat. Neurosci.* **2**, 94–98 (1999).

258. Tuerk, P. W. *et al.* Augmenting treatment efficiency in exposure therapy for PTSD: a randomized double-blind placebo-controlled trial of yohimbine HCl. *Cogn. Behav. Ther.* **47**, 351–371 (2018).

259. Rosa, J., Myskiw, J. C., Furini, C. R. G., Sapiras, G. G. & Izquierdo, I. Fear extinction can be made state-dependent on peripheral epinephrine: Role of norepinephrine in the nucleus tractus solitarius. *Neurobiol. Learn. Mem.* **113**, 55–61 (2014).

260. Mueller, D., Porter, J. T. & Quirk, G. J. Noradrenergic signaling in infralimbic cortex increases cell excitability and strengthens memory for fear extinction. *J. Neurosci.* **28**, 369–75 (2008).

261. Follesa, P. *et al.* Vagus nerve stimulation increases norepinephrine concentration and the gene expression of BDNF and bFGF in the rat brain. *Brain Res.* **1179**, 28–34 (2007).

262. Berlau, D. J. & McGaugh, J. L. Enhancement of extinction memory consolidation: the role of the noradrenergic and GABAergic systems within the basolateral amygdala. *Neurobiol. Learn. Mem.* **86**, 123–32 (2006).

263. Robertson, S. D., Plummer, N. W., de Marchena, J. & Jensen, P. Developmental origins of central norepinephrine neuron diversity. *Nat. Neurosci.* **16**, 1016–1023 (2013).

264. George, M. S. *et al.* A pilot study of vagus nerve stimulation (VNS) for treatment-resistant anxiety disorders. *Brain Stimul.* **1**, 112–121 (2008).

265. Cristancho, P., Cristancho, M. A., Baltuch, G. H., Thase, M. E. & O'Reardon, J. P. Effectiveness and safety of vagus nerve stimulation for severe treatment-resistant major depression in clinical practice after FDA approval: outcomes at 1 year. *J. Clin. Psychiatry* **72**, 1376–82 (2011).

266. Roelofs, K. Freeze for action: Neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 372 (2017).

267. Sawatari, H. *et al.* Identification and characterization of an insular auditory field in mice. *Eur. J. Neurosci.* **34**, 1944–52 (2011).

268. A, K., T, D., K, O., H, I. & Y, T. Efferent Connections of the Ventral Auditory Area in the Rat Cortex: Implications for Auditory Processing Related to Emotion. *Eur. J. Neurosci.* **25**, (2007).

269. Hersman, S., Allen, D., Hashimoto, M., Brito, S. I. & Anthony, T. E. Stimulus salience determines defensive behaviors elicited by aversively conditioned serial compound auditory stimuli. *Elife* **9**, (2020).

270. Wotjak, C. T. Sound check, stage design and screen plot – how to increase the comparability of fear conditioning and fear extinction experiments. *Psychopharmacology (Berl).* **236**,.

271. Fadok, J. P. *et al.* A competitive inhibitory circuit for selection of active and passive fear responses. *Nature* **542**, 96–100 (2017).

272. Barthó, P. *et al.* Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J. Neurophysiol.* **92**, 600–8 (2004).

273. Roux, L., Stark, E., Sjulson, L. & Buzsáki, G. In vivo optogenetic identification and manipulation of GABAergic interneuron subtypes. *Curr. Opin. Neurobiol.* **26**, 88–95 (2014).

274. Wiegert, J. S., Mahn, M., Prigge, M., Printz, Y. & Yizhar, O. Silencing Neurons: Tools, Applications, and Experimental Constraints. *Neuron* **95**, 504–529 (2017).

275. Lonsdorf, T. B. *et al.* Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neurosci. Biobehav. Rev.* **77**, 247–285 (2017).

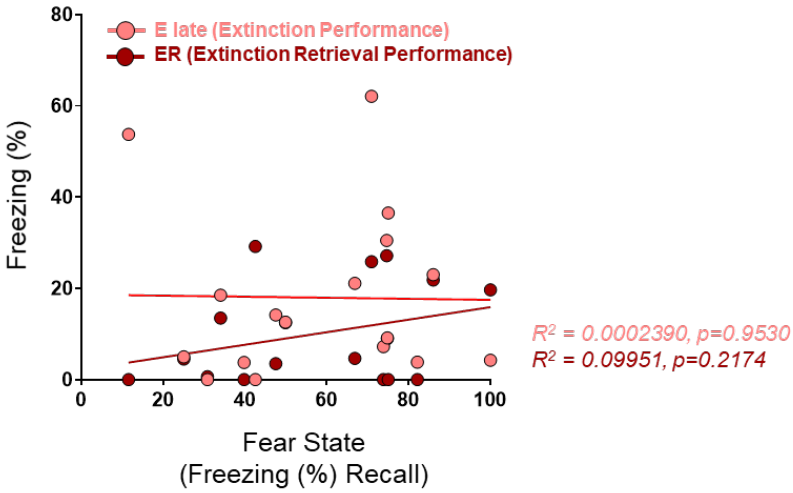
276. Kamprath, K. & Wotjak, C. T. Nonassociative learning processes determine expression and extinction of conditioned fear in mice. *Learn. Mem.* **11**, 770–86 (2004).

277. McNaughton, N. & Corr, P. J. A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neurosci. Biobehav. Rev.* **28**, 285–305 (2004).

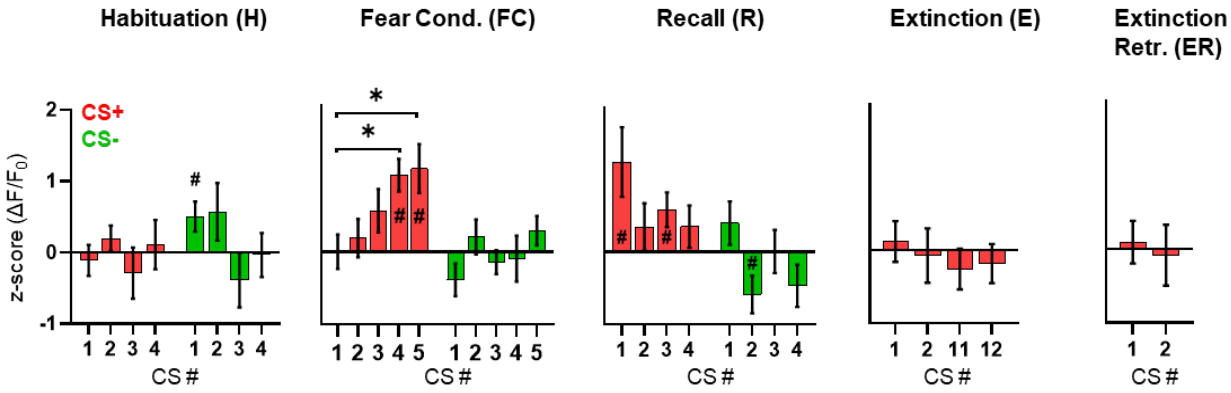
278. Perusini, J. N. & Fanselow, M. S. Neurobehavioral perspectives on the distinction between fear and anxiety. *Learn. Mem.* **22**, 417–25 (2015).

279. Dolensek, N., Gehrlach, D. A., Klein, A. S. & Gogolla, N. Facial expressions of emotion states and their neuronal correlates in mice. *Science* **368**, 89–94 (2020).

5.3 Supplementary figures

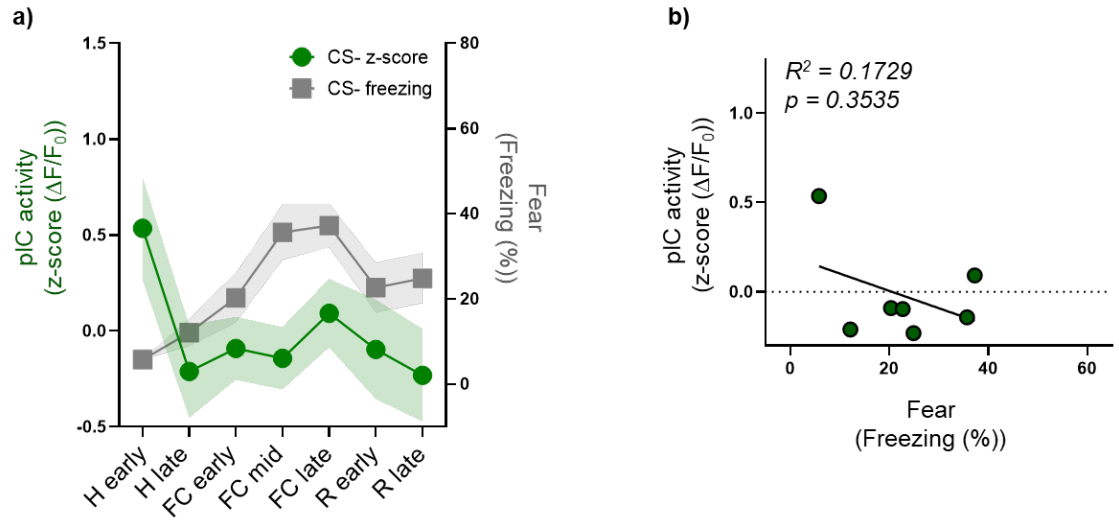


Supplementary figure 1: Extinction performance and retrieval do not depend on initial fear learning (related to Fig. 1).
Correlation between fear state (average % freezing to all four CS+ during recall (R)) and extinction performance (average % time freezing during extinction late (E late)) or extinction retrieval performance (average % time spent freezing during extinction retrieval (ER)). Linear regression reveals that the fear state cannot predict extinction performance (N=17 mice, F(1,15)=0.003586, R^2=0.0002390, slope not different from zero, p=0.9530), or extinction retrieval (N=17 mice, F(1,15)=0.09951, R^2=0.06052, slope not different from zero, p=0.2174).



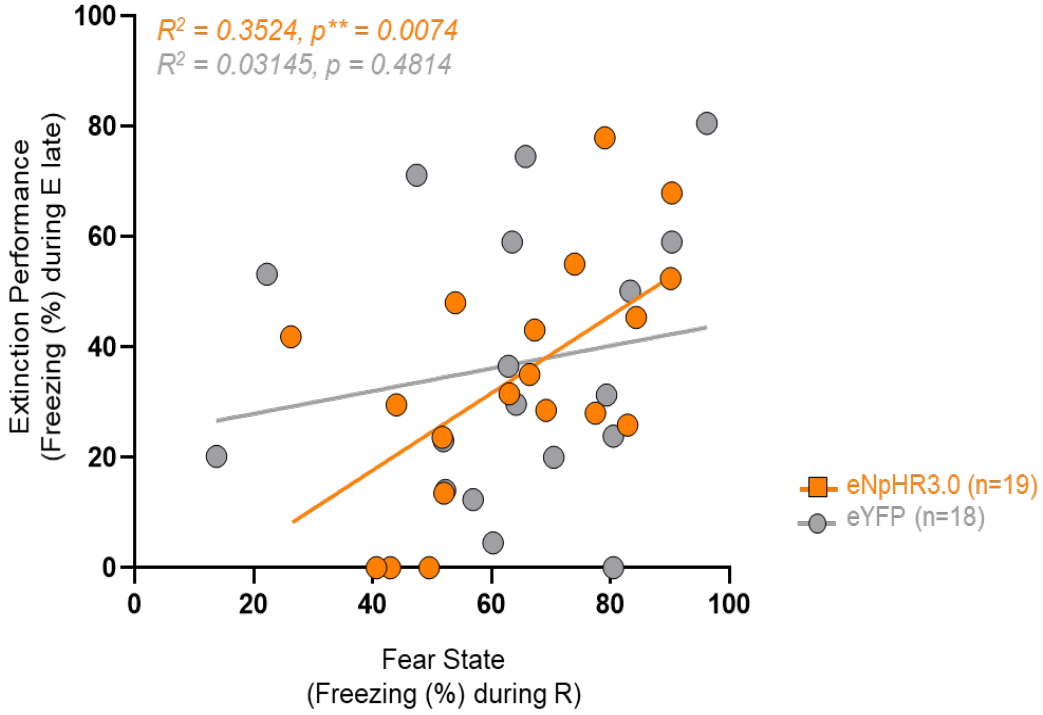
Supplementary figure 2: pIC activity is gradually modulated by single fear-eliciting cues within sessions (related to Fig. 3).

*Average z-scored pIC responses to each CS+ presentation during habituation (H), fear conditioning (FC) and recall (R), and extinction retrieval (as well as to first and last two CS+ during extinction (E). One-sample t test with zero as theoretical mean show that pIC excitatory neuron activity is significantly increased by CS+4 and 5 during FC (CS+4: t=4.747, df=17, $p^{###}$=0.0002, CS+5: t=3.418, df=17, $p^{##}$=0.0033) as well as by CS+ 1 and CS+3 during recall (CS+1: t=2.589, df=17, $p^{#}$=0.0191, CS+3: t=2.435, df=17, $p^{#}$=0.0262). CS- evoked activity is significantly activated by the first CS- during habituation (CS-1: t=2.384, df=16, $p^{#}$=0.0298), but interestingly, it is significantly deactivated by CS-2 during recall (t=2.271, df =17, $p^{#}$=0.0364). I used a one-way RM ANOVA to compare CS evoked activity within each session, which revealed a significant increase in CS+ evoked activity during FC (N=18 mice, F(3.580, 60.87)=4.055, p**=0.0073; Bonferroni corrected* post hoc *tests indicate a significant difference between CS+1 vs CS+4, p*=0.0430 and between CS+1 vs CS+5, p*=0.0206) as well as a significant decrease of CS- evoked activity during recall (N=18 mice, F(3,51)=3.133, p*=0.0334, Bonferroni corrected* post hoc *test reveal no significant difference between each CS-). Data are plotted as mean ± SEM.*

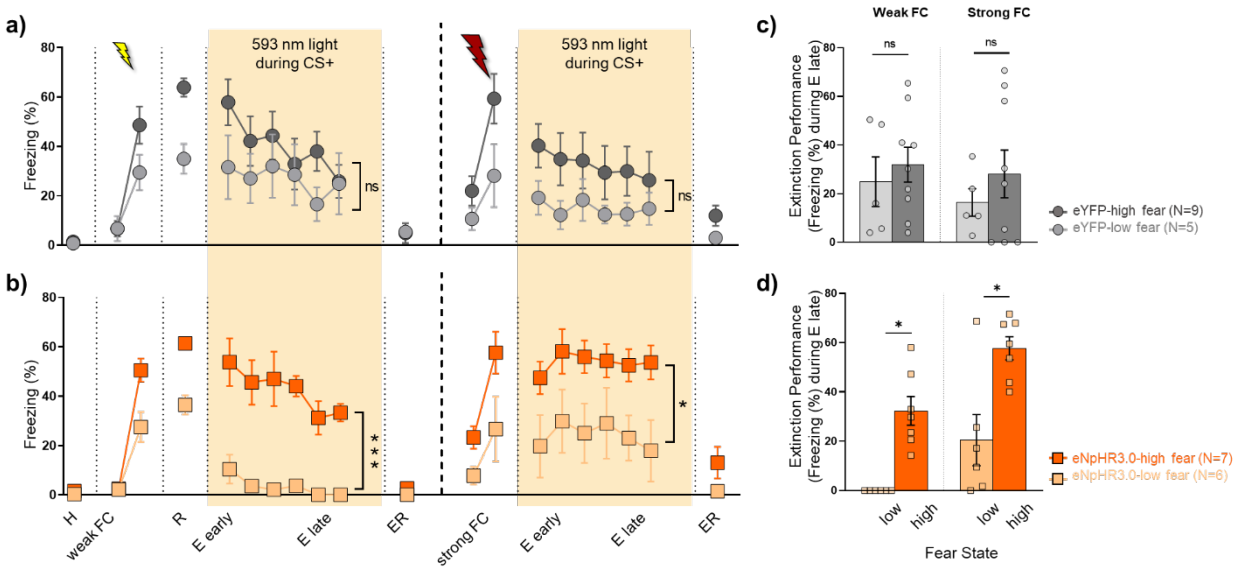


Supplementary figure 3: CS- evoked pIC activity does not correlate with fear levels (related to Fig. 5).

a) *CS- evoked pIC excitatory neurons activity is shown as average z-score or % freezing during for two consecutive CS- presentations (except for FC mid, where only one CS+ is shown). CS- evoked pIC activity is stable across the sessions of the paradigm. N=18 mice, data are plotted as mean ± SEM.* ***b)*** *Same data as in* ***a)*** *plotted using a linear regression showing no correlation between CS- evoked fear behavior and pIC activity (N=18 mice, F(1,5)=1.045, R^2=0.1729, slope significantly non-zero, p=0.3535).*

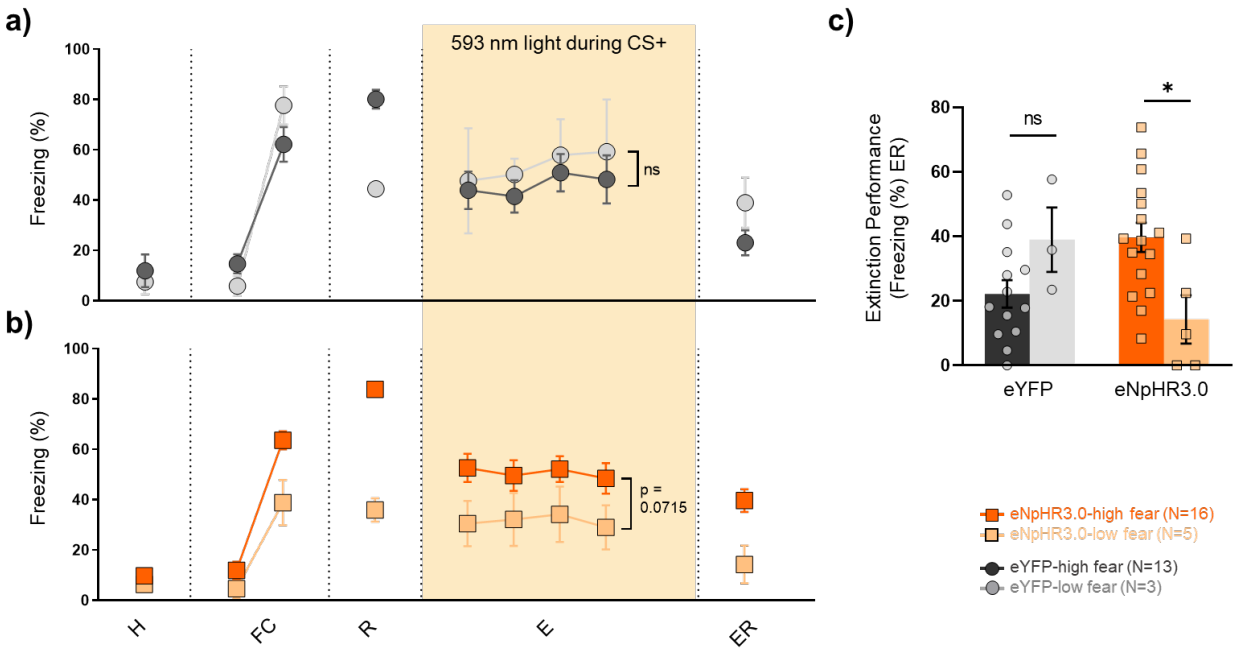


Supplementary figure 4: Internal fear state predicts extinction performance in eNpHR3.0 animals (related to Fig. 15 and 16).
*Correlation between 'fear state' (% time spent freezing during recall (R)) and 'extinction performance' (% time spent freezing during extinction late (E late)) pooled together from both groups (weak and strong FC). Linear regression reveals a strong positive correlation in eNpHR3.0 animals (F(1,17)=9.251, R^2=0.3524, slope significantly non-zero, p**=0.0074), whereas in eYFP controls, the correlation is much weaker (F(1,16)=0.5196, R^2=0.03145, slope not different from zero, p=0.4814.*

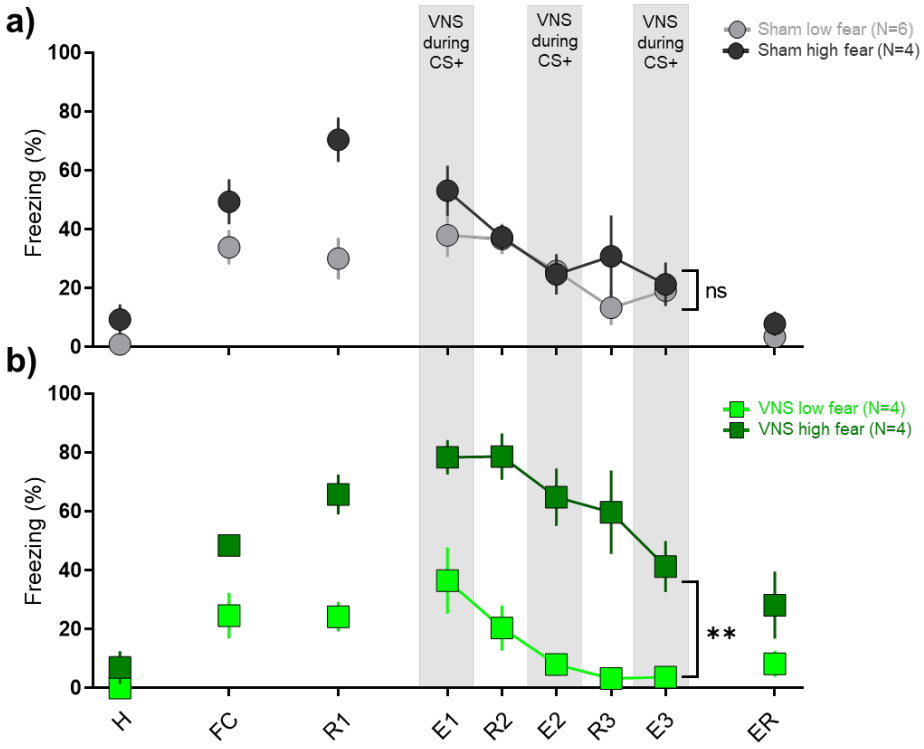


Supplementary figure 5: State-dependent effects of optogenetic pIC inhibition on extinction performance (related to Fig. 18).

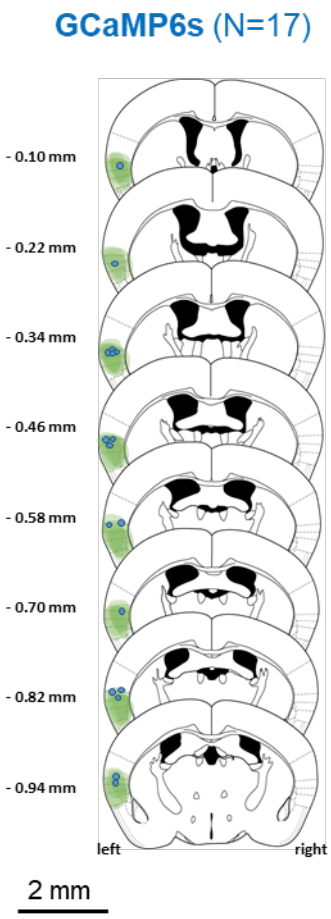
a) *and* ***b)*** *CS+ evoked freezing across all stages of the FC and Ext paradigm (same as in* ***Fig. 17 c)*** *and* ***18 b)*** *), but high and low fear animals compared within the opsin groups.* ***a)*** *High fear state eYFP controls extinguish at a similar rate as the low fear state eYFP controls both after weak and after strong FC (N=9 eYFP-high fear and N=5 eYFP-low fear, two-way RM ANOVAs; extinction after weak FC: group (fear state) effect, F(1, 13)=0.9825, p=0.3397; time effect, F(5, 65)=6.142, p***=0.0001, group x time interaction F(5,65)=2.772, p*=0.0249; extinction after strong FC: group (fear state) effect, F(1, 13)=2.018, p=0.1790; time effect, F(5, 65)=1.045, p=0.3990, group x time interaction F(5,65)=0.3005, p=0.9109).* ***b)*** *In contrast, high and low fear eNpHR3.0 animals extinguish at significantly different rates, both after weak and after strong FC (N=7 eNpHR3.0-high fear and N=6 eNpHR3.0-low fear, two-way RM ANOVAs; extinction after weak FC: group (fear state) effect, F(1, 12)=25.48, p***=0.0003; time effect, F(2.756, 33.07)=4.071, p*=0.0166, group x time interaction F(5,60)=0.8763, p=0.5025; extinction after strong FC: group (fear state) effect, F(1, 12)=5.603, p*=0.0365; time effect, F(2.975, 35.70)=2.065, p=0.1227, group x time interaction F(5,60)=0.4333, p=0.8236).* ***c)*** *and* ***d)*** *Comparison of extinction performances between high and low fear state animals within opsin groups, measured as average % freezing during extinction late.* ***c)*** *There is no difference in extinction performance between eYFP high or low fear animals (two-tailed unpaired t tests; weak FC: t=0.5720, df=12, p=0.5779; strong FC: t=0.8361, df=12, p=0.4194).* ***d)*** *In contrast, there is a significant difference in extinction performance between eNpHR3.0 high and low fear animals (two-tailed unpaired t tests; weak FC: t=5.103, df=11, p***=0.0003; strong FC: t=3.444, df=11, p=0.0055). All data are plotted as mean ± SEM.*



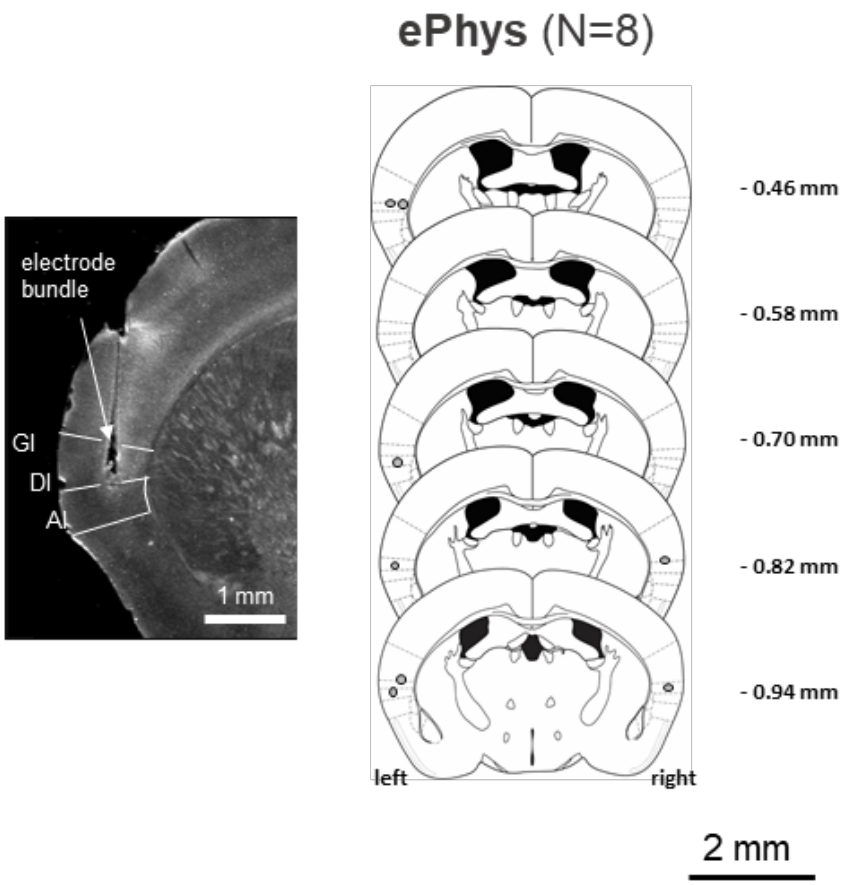
Supplementary figure 6: State-dependent effects of pIC silencing on extinction memory retrieval (related to Fig. 19).
a) *and* ***b)*** *CS+ evoked freezing across all stages of the FC/ Ext paradigm (same as in* ***Fig. 19 e)*** *), but high and low fear animals compared within the opsin groups.* ***a)*** *High fear state eYFP controls extinguish at a similar rate as the low fear state eYFP controls (N=13 eYFP-high fear and N=3 eYFP-low fear, two-way RM ANOVAs of extinction; group (fear state) effect, F(1, 11)=0.2702, p=0.6135; time effect, F(2.347, 25.82)=1.013, p=0.3877, group x time interaction F(3,33)=0.1030, p=0.9578).* ***b)*** *In contrast, high and low fear eNpHR3.0 animals extinguish at close to significantly different rates (N=16 eNpHR3.0-high fear and N=5 eNpHR3.0-low fear, two-way RM ANOVAs of extinction; group (fear state) effect, F(1, 19)=3.645, p=0.0715; time effect, F(2.660, 50.54)=0.2932, p=0.8066, group x time interaction F(3,57)=0.09450, p=0.9628).* ***c)*** *Comparison of extinction retrieval between high and low fear animals within opsin groups, measured as average % freezing during extinction late. There is no difference in extinction performance between eYFP high or low fear animals (two-tailed unpaired t test; t=1.673, df=14, p=0.1165). In contrast, there is a significant difference in extinction retrieval between eNpHR3.0 high and low fear animals (two-tailed unpaired t test; t=2.779, df=19, p*=0.0120). All data are plotted as mean ± SEM.*



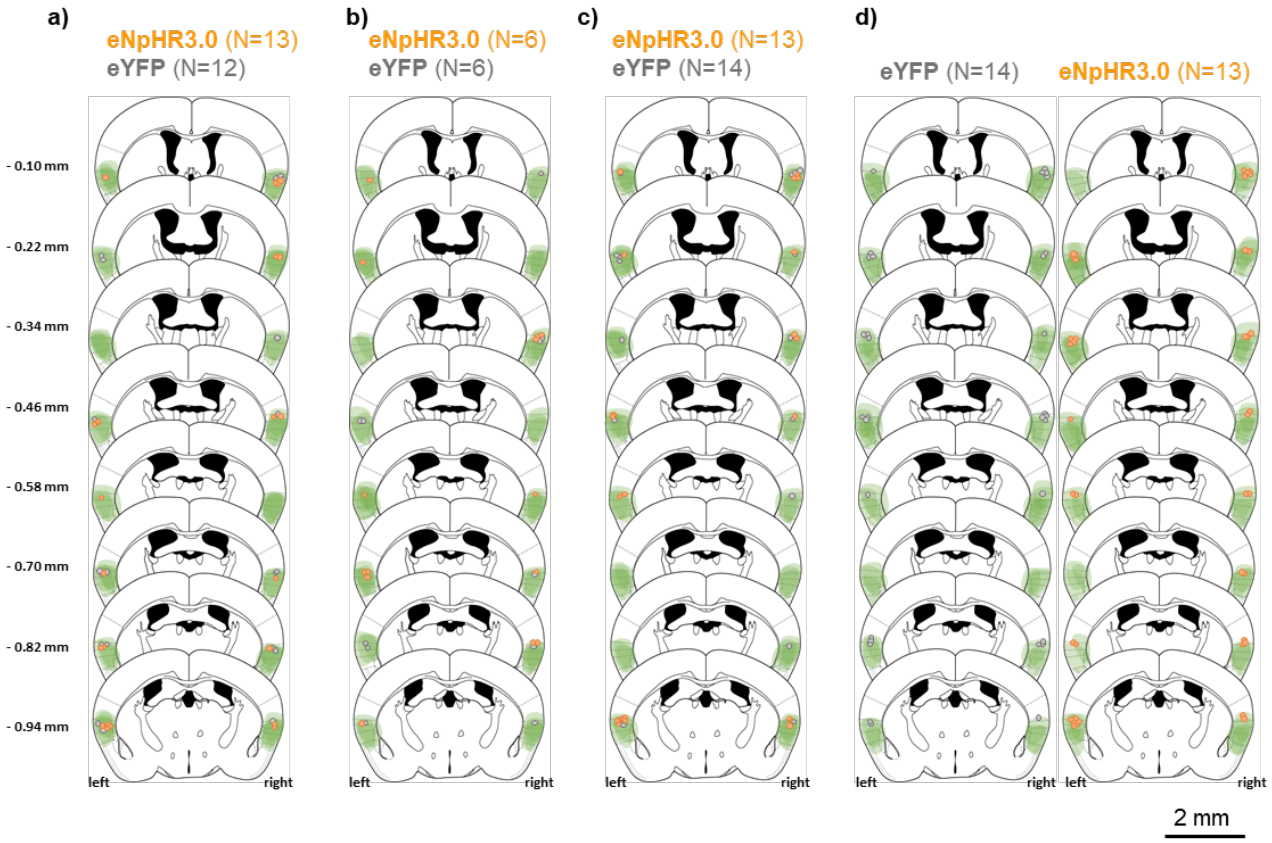
Supplementary figure 7: Vagus nerve stimulation (VNS) renders fear extinction state-dependent (related to Fig. 20).
a) *High and low fear state Sham controls extinguish at a similar rate, irrespective of the fear state (N=6 Sham high fear and N=4 Sham low fear, two-way RM ANOVAs during extinction days (extinction 1 (E1) to extinction 3 (E3)); group (fear state) effect, F(1, 8)=1.080, p=0.3291; time effect, F(3.128, 25.03)=7.085, p**=0.0012, group x time interaction F(4, 32)=1.1640, p=0.3449).* ***b)*** *High fear state VNS animals show significantly higher freezing values throughout the extinction days (extinction 1 (E1) to extinction 3 (E3)) in comparison to low fear state VNS animals (N=4 VNS high fear and N=4 VNS low fear, two-way RM ANOVAs during extinction days (extinction 1 (E1) to extinction 3 (E3)); group (fear state) effect, F(1, 6)=33.05, p**=0.0012; time effect, F(1.912, 11.47)=11.22, p**=0.0022, group x time interaction F(4,24)=1.336, p=0.2854).*



Supplementary figure 8: Histological verification of injection- and implantation sites for mice used in fiber photometry experiments (related to Fig. 2 to 6).
Schematic visualization of GCaMP6s expression (green-shaded regions). Blue dots represent the center of tip of the implanted optic fiber.

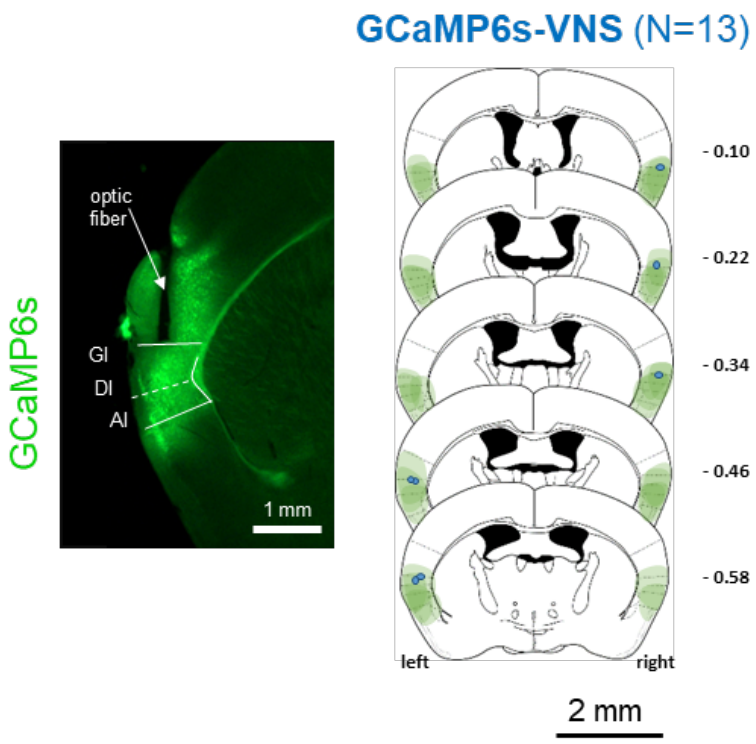


Supplementary figure 9: Histological verification of electrode implantation sites for mice used in single unit electrophysiology experiments (related to Fig. 8 to 13).
Left: Example histological image for an electrode implantation site in pIC. Right: Grey dots represent the center of the longest tip of the electrode bundle.



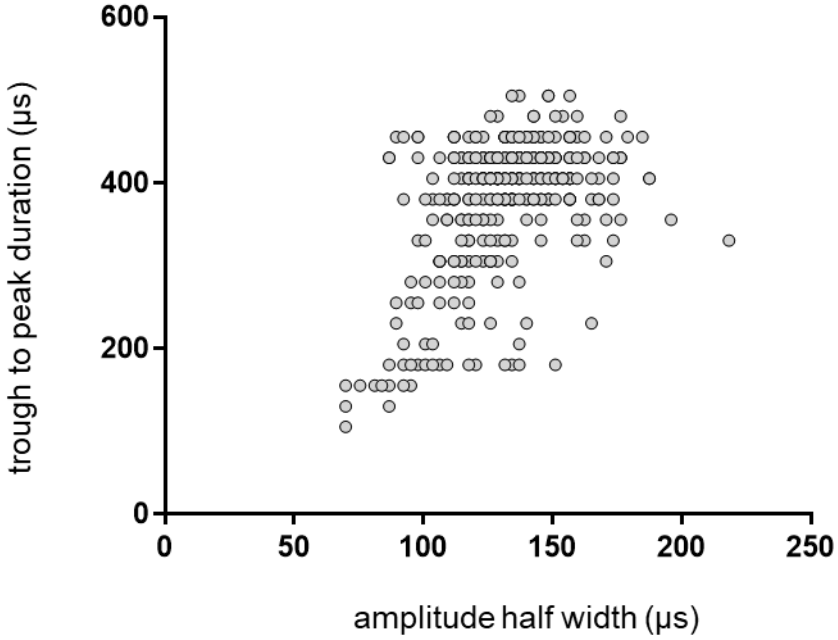
Supplementary figure 10: Histological verification of injection- and implantation sites for eNpHR3.0 and eYFP mice (related to Fig. 15 to 19).

Schematic visualization of eNpHR3.0 or eYFP expression (green-shaded regions). Colored dots (orange: eNpHR3.0, grey: EYFP) represent the center of tip of the implanted optic fibers for eNpHR3.0 or eYFP expressing animals, respectively. ***a)*** *Animals used for figure 15. Note that anatomy from N=3 eNpHR3.0 and N=5 eYFP was not available.* ***b)*** *Animals used for figure 16.* ***c)*** *Animals used for figure 17 and 18. Note that anatomy from N=2 eNpHR3.0 and N=5 eYFP was not available.* ***d)*** *Animals used for figure 19.*



Supplementary figure 11: Histological verification of injection- and implantation sites for GCaMP6s-VNS mice (related to Fig. 20 and 21).

Left: Example histological image for a virus expression (green) and optic fiber implantation site in pIC. Right: Schematic visualization off GCaMP6s expression (green-shaded regions). Blue dots represent the center of tip of the implanted optic fibers. Note that histological data were only available for 7 out of 13 mice.



Supplementary figure 12: Cell class identification using waveform analysis (related to Fig. 13 to 18). *To identify neuronal subtypes (i.e. pyramidal cells vs interneurons), amplitude half width and trough to peak duration of their extracellular waveforms were analyzed. Fast spiking interneurons and excitatory pyramidal cells should fall into two separate clusters, but cluster separation is very poor here, probably due to a limited cell number.*