Cellular and Oscillatory Substrates of Extinction Learning

A thesis submitted by

Patrick Davis

in partial fulfillment of the requirement for the degree of

PhD

in

Neuroscience

Tufts University

Sackler School of Graduate Biomedical Sciences

May 2017

Adviser: Leon Reijmers, PhD

Abstract

In contrast to the hard-wired components characteristic of electronic circuits and some invertebrate nervous systems, mammalian brains contain modifiable elements within circuits that are capable of both robustness and remarkable adaptability. This balance is exemplified by the capacity of mammals to both learn and update previously formed associations, and to select from opposing behavioral strategies based on a diversity of previous experience. A classic example of this is the paradigm of contextual fear conditioning and extinction learning, whereby an animal first forms a fearful association with a conditioned context, but subsequently learns that the context is safe. Because these two learning processes have distinct and opposing behavioral consequences, this paradigm grants experimental access to the features of the brain which enable both the formation of an associative memory and the subsequent modification of that memory with continued experience, as well as to the circuits which govern ultimate behavioral output. To date, the mechanisms underlying the modification of associative fear memories through extinction learning have yet to be completely understood. Using a novel combination of chemo- and optogenetics, activity-based neuronal-ensemble labeling, circuit tracing, and *in vivo* electrophysiology, we have identified cellular and oscillatory substrates of fear extinction learning that depend critically on parvalbumin (PV)- expressing interneurons in the basolateral amygdala (BLA). Specifically, we found that extinction learning confers PV- interneurons in the BLA with a dedicated role in the

suppression of a previously encoded fear memory via selective suppression of BLA fear-encoding neurons. BLA PV-interneurons are positioned to gate reciprocal communication between BLA and medial prefrontal cortex (mPFC), and their activity controls the activation of spatially localized and functionally opposing ensembles within mPFC. We establish that BLA PV-interneurons are critical to the generation of two opposing oscillatory states across the mPFC-BLA circuit. Extinction learning modifies the relationship between these two states to allow for "competition" between them. Artificial induction of these oscillatory states by direct manipulation of PV-interneurons is sufficient to elicit bidirectional, learning-dependent control over mPFC-BLA circuit coordination and fear behavior. Finally, we provide evidence potentially linking these circuit oscillatory properties with ensemble data through the phenomenon of resonance. These findings identify cellular and oscillatory substrates of fear extinction learning that critically depend on BLA PV-interneurons. The role of the BLA PV-network in mediating interactions between previously learned functional network states and memory-encoding ensembles is likely to be broadly applicable to interneuron microcircuit function in both physiological and pathological states.

Dedication

To my Mom and Dad, who taught me the value of my own curiosity, and to Alex, who taught me how to use it.

Acknowledgments

Leon Reijmers, PhD

Alexandros Poulopoulos, PhD

Jeffrey Macklis, MD, HST

Hari Padmanabhan, PhD

Jenny Sasaki, PhD

Minagi Ozawa

Alex Jones

Joe Zaki

Steffani Viola

Peter Gibb

Gabrielle Girardeau, PhD

Thomas Papouin, PhD

Jamie Maguire, PhD

Drew Hooper, PhD

Semion Salkin, PhD

Table of Contents

List of Figures

List of Abbreviations

BLA: basolateral amygdala

PV: parvalbumin

PV-interneuron: parvalbumin-expressing interneuron

mPFC: medial prefrontal cortex

PLC: prelimbic cortex, subdivision of mPFC

ILC: inframlimbic cortex, subdivision of mPFC

CeA: central amygdala

CA1: cornu ammonis area 1 of hippocampus

CS: conditioned stimulus

US: unconditioned stimulus

LTP: long-term potentiation

STDP: spike-time-dependent plasticity

NMDA: N-methyl-D-aspartate

PTSD: Post-traumatic stress disorder

PPE: positive predictive error

IEG: immediate early gene

Chr2: Channelrhodopsin-2

IPSP: inhibitory post-synaptic potential

EEG: electroencephalogram

LFP: local field potential

DREADD: Designer Receptor Exclusively Activated by Designer Drug

GFP: green fluorescent protein

ZIF: Zif268, IEG

CREB: cAMP response element binding protein, IEG

FC: fear conditioning

EXT: extinction

HC: homecage

NS: no-shock

CtxB: context-B

PSD: post-synaptic density

Chapter 1: Introduction

One of the fundamental tenets of modern neuroscientific thinking is that arrays of individual synaptic and neuronal components underlie an essentially infinite and arbitrary set of complex computations leading to a discrete set of output behaviors. As such, neuronal circuits require a delicate balance of robustness and plasticity in response to sensory inputs, such that they can facilitate both acutely appropriate and adaptable behavior. One of the hallmarks of such circuits is that they exhibit complex properties across multiple emergent levels, and that the properties of the emergent levels influence their substituent components at lower levels. In this thesis, I have endeavored to address, using the classical conditioning model of associative fear and extinction learning, how emergent circuit properties are produced -- and, importantly, modified by learning -- to govern both the actions of individual neuronal components and, ultimately, the behavior of the animal.

Learning and Memory

The operational perspective of the brain defines memory as any experience-dependent alteration in neural function. Broadly, memory can be sub-divided into two categories, explicit and implicit, a classification schema that, although it was first introduced by psychologists years before, was made famous by Patient HM, who had complete anterograde explicit memory amnesia, but retained the ability to learn complex procedural (implicit) tasks [1,2]. Explicit memory, in its most basic definition, is the subjective experience of memory, and thus requires conscious thought. The ten digits of one's phone number or the name and face of one's first love are both clear examples of

information that is necessarily encoded in explicit memory. They can be recalled volitionally, and they require conscious thought; in fact, they *are* the conscious thought. The subjective experience of recalling a name or a face are experience dependent, conscious recollection events. In contrast, implicit memory can be loosely defined as a change in the sensory input – motor output relationship of the brain, and need not invoke conscious recall. In other words, implicit memories are experience-dependent changes that alter the production of behavior in response to, or depending on, a given set of sensory stimuli. Because they do not require access to the subjective experience of memory (ie: consciousness), implicit memories can be studies in non-human animals, including both invertebrates and vertebrates. The past half-century has seen massive gains in the understanding of the neural mechanisms of memory formation and retrieval, owing almost entirely to the study of implicit memories [3].

Implicit memory includes several more specific forms of learning, including procedural and motor-based learning. The form of implicit memory most commonly studied is known as classical conditioning, whereby an animal's response to a neutral stimulus (conditioned stimulus, CS) can be altered by pairing, or associating it, with a stimulus which produces an innate behavioral and neural response (unconditioned stimulus, US) [4,5]. This form of learning was described and popularized by Pavlov, who demonstrated that animals could be "taught" to exhibit a salivary response (conditioned response) to the sound of a bell ringing (CS) if it was paired consistently with the smell of food (US). Because implicit learning, in particular associative or classical conditioning, results in highly reproducible and discrete alterations in the behavior of

animals, it can be studied directly, and the molecular, cellular, and neural circuit components which are necessary for it have been interrogated extensively.

Throughout the history of early modern neuroscience, two theories dominated the discussion among neuroscientists of memory encoding and formation within the brain. The "connectionist theory," put forth by Cajal and advanced by many, including Hebb, Morris, and Kandel, after him, holds that memories are encoded in alterations in the connections between neurons [6–8]. An alternative proposal, put forth by Lashley and known as the "aggregate field theory", argued against localized and discrete functions for individual neurons or connections between neurons, instead placing emphasis on the collective bioelectric field generated by many neurons and even many brain regions [9]. Though this idea has fallen out of favor, evidence from both animal models, using direct electrical recordings, and human subjects, using fMRI, indicate that memory storage and retrieval frequently relies on distributed neural activity throughout multiple brain regions and involving large populations of neurons. I will address specific evidence for the role of aggregate electric activity, in particular oscillatory activity, in neural function later in this introductory chapter. Suffice it to say, however, the focus of modern neuroscience has shifted drastically toward the connectionist theory, owing largely to 1) pioneering work in invertebrate models of implicit memory and 2) the parallel discoveries of mechanisms of synaptic plasticity in mammalian circuits. Excellent reviews of both subjects can be easily found, but I will briefly review the work as it pertains to general learning theory here.

The first concrete evidence for synaptic plasticity – in other words, experience-dependent changes in synaptic strength – as an underlying feature of learning and

memory came from investigation of the invertebrate sea snail *Aplysia Californica*, which exhibits a characteristic gill-withdrawal reflex upon sensory stimulation of the siphon organ. Early pioneering work demonstrated that the gill-withdrawal reflex was subject to both habituation, upon repeated stimulation, and sensitization, upon pairing with an electric shock [10]. These rudimentary forms of learning were localized to changes in the strength of connections between a small number of neurons in the reflex arc. In addition, continued work demonstrated that the learning was dependent on the neuromodulator serotonin, and the downstream biochemical mechanisms have since been largely elucidated. A partial list of these includes G-protein coupled receptor activation, adenyl cyclase activity, cyclic-AMP signaling, modification of post-synaptic receptors, and activation of cAMP-response-element-binding protein (CREB) [10]. CREB is of particular note because it, for the first time, linked activity of synapses to changes in nuclear signaling and transcriptional responses. This work, in total, was extremely important for a number of reasons: 1) it established that simple forms of implicit learning could be studied in invertebrate models, 2) it established that certain forms of memory could be attributed to modifications of individual synapses, 3) it localized those synapses to the reflex arc which generated the innate behavior, suggesting the memory could be due to changes in synaptic weight among neuron *within* the circuit responsible for a given behavior (arguing against the existence of dedicated "memory" cells), and 4) it began the investigation into the molecular pathways that are thought to underlie memory formation. Indeed, many of the biochemical pathways found in Aplysia have directly analogous functions in the mammalian brain, demonstrating the influence and importance of this pioneering work in *Aplysia* [8].

A second, parallel line of breakthrough discoveries further supported the connectionist approach. The discovery of long-term potentiation (LTP) – a robust and experimentally inducible increase in the strength of pathway-specific glutamatergic synapses -- in rabbit hippocampal slice preparations provided the phenomenological leap from invertebrate to mammalian brains [11,12]. After all, if changes in the strength of single synapses can underlie memory formation in *Aplysia*, why not in more complex systems as well? From the pioneering work of Lomo, Andersen and colleagues we know that changes in synaptic strength can be reproducibly effected in hippocampal slices. Moreover, continued work over the past several decades has conclusively demonstrated that the molecular requirements for these changes are also required for the acquisition or stabilization of many forms of memory. A full review of the molecular mechanisms of synaptic plasticity is far beyond the scope of this introduction, but I will call attention to one example as it sheds light on current models of memory formation. For example, 100Hz stimulation of performant path leads to classical LTP, and much elegant work has demonstrated that this form of plasticity is dependent on NMDA receptor activation and function [13]. Genetic and pharmacologic evidence has conclusively demonstrated that disruption of NMDAR function can perturb both *ex vivo* plasticity (eg: LTP) as well as hippocampal-dependent learning, such as spatial memory tasks, further bolstering the argument that synaptic plasticity underlies memory formation, and consistent with their known role in coincidence detection [14,15]. In addition, much important work has elucidated additional biochemical components, including kinases (CaMKII), scaffolding molecules (PSD95), neurotrophic signaling molecules (BDNF), and nuclear signaling complexes (CREB, IEGs) that are essential for both canonical synaptic plasticity and for

many types of learning [16]. With this evidence in mind, Kandel wrote, in 1998, "Current models of brain function hold that learning corresponds to changes in the efficacy of single synapses" [17]. Some important caveats should temper our excitement about having solved the puzzle of memory formation. While the evidence is conclusive that synaptic plasticity plays a role in memory formation, it is not clear whether these modifications are stable throughout the duration of the memory (often the lifetime of the animal or human subject). In other words, does synaptic plasticity facilitate memory formation, or *is it* the memory itself? Moreover, it is not clear what role synaptic plasticity plays in shaping the aggregate activity of neuronal circuits, a topic I will review later in this chapter and in the Discussion chapter. Any fully formed theory of memory formation will have to take into account both reductionist evidence from simplified model organisms and memory tasks as well as the evidence that mammalian, in particular human, brains exhibit properties and behaviors that lie outside the explanatory power of simplified wiring diagrams with modifiable connections. Such a theory would necessitate work in more complex brains, with sophisticated learning paradigms and experimental access to the synaptic, cellular and systems (emergent) levels of function.

Fear Conditioning and Extinction Learning

Fear conditioning is form of classical conditioning whereby an animal learns to associate a previously neutral stimulus or context with a noxious stimulus, thereafter exhibiting fearful physiologic behavior in the presence of the previously neutral stimulus [18]. This most commonly accomplished by exposing an animal to a novel context paired with a series of footshocks, after which the animal becomes immobile ("freezes") during

subsequent exposures to the conditioned context. At the synaptic level, classical studies demonstrated that this form is learning is dependent on NMDARs within a structure known as the basolateral amygdala [19–21]. Due to the canonical role of NMDARs in coincidence-detection and synaptic plasticity, these findings led to the prevailing theory that a primary substrate of associative fear learning (and indeed associative learning in general) results from spike-time-dependent-plasticity within the BLA [16,22,23]. This theory holds that sub-threshold input carrying information about the conditioned stimulus (context) is, when paired with input carrying footshock and pain information (US), transformed to supra-threshold by NMDAR-dependent STDP, thereby making CS presentation sufficient to elicit amygdalar responses thereafter. This model is supported by the fact that BLA receives dense innervation from thalamus, which encodes painful stimuli, and hippocampus, which is thought to encode contextual information [24–26]. Indeed, single unit recordings in BLA have demonstrated that putative pyramidal neurons within BLA can respond to footshock, indicating that the US information is natively supra-theshold and could therefore convert paired CS information to suprathreshold via STDP [26–29]. The most direct support of this model, however, came from a recent study which was able to artificially create an associative fear memory, or at least its behavioral correlate, by bi-directional tuning of STDP at CS-US synapses within the BLA [30]. Thus, it is widely accepted that the amygdala plays a critical role in forming the associations which, in part, underlie conditioned fear memories. However, much is still unknown about the precise function of BLA fear-encoding neurons, or the BLA in general, namely 1) what features do BLA neurons encode, and which are consistent across the multiple types of BLA-dependent learning (ie: fear, extinction, reward) and 2) how does the

connectivity, in particular the output of BLA principal neurons, inform their function? The latter question has begun to be addressed with several elegant studies, but mapping principal neuron function onto hodology is fraught with equivocal experimental results for several reasons, most obvious of which is that functionally opposed groups ("ensembles") of neurons appear to project to largely overlapping brain regions [31–36]. A more sophisticated understanding of the downstream effects of BLA neuron activation, as well as the parameters controlling interpretation in downstream regions, is required to resolve this dilemma. As for the former question, a debate continues regarding the precise set of conditions that lead to BLA activation, and what "features" their activity encodes. Though the answer of "valence specific associations" has gained popularity in recent years, alternative theories posit that the amygdala encodes emotional salience and/or prediction errors [37–40]. The particular theory one favors is highly correlated with technique and behavioral output used.

The amygdala is also critical for a separate form of learning known as fear extinction, whereby a previously conditioned animal is repeated exposed to the conditioned stimulus, resulting in a diminished physiologic fear response[25]. This form of learning is analogous to exposure therapy, a commonly used treatment for human patients with post-traumatic stress disorder (PTSD) and other anxiety disorders [42,43]. As in humans, extinction therapy in rodents results in a transient reduction in fear behavior, and conditioned fear responses often return spontaneously with the passage of time (recovery) or exposure to an additional stressor (termed "reinstatement"). This critical behavioral phenomenon is enlightening for two reasons: 1) a better understanding of extinction learning is necessary to develop more robust treatments for human patients, which are

currently limited by spontaneous return of fear, and 2) extinction learning does not merely erase the previously learned fear association; rather, it creates an additional overlaid memory which is capable of suppressing conditioned fear responses. In addition, studies from extinction of cue-conditioned rodents have demonstrated that the extinction-induced fear suppression is specific to the context in which it is learned, again analogous to the experience of patients treated with exposure therapy, and indicating the presence of a new and discrete context-based memory formation that suppresses the original fear memory [44,45].

Classical lesion studies demonstrated the necessity of a properly functioning amygdala for both the encoding and retrieval of conditioned fear responses, as well as for extinction learning [46]. Similar studies have also implicated several other brain regions as critical for fear conditioning, including the central amygdala (CeA), medial prefrontal cortex (mPFC), hippocampus, bed-nucleus of the stria-terminalis (BNST), and several others [47]. The distributed nature of such learning and behavioral phenomena raises the obvious question of what, precisely, each of these regions contributes. In addition, standard approaches to circuit dissection are inherently confounded by the densely and reciprocally interconnected nature of neuronal circuits, allowing the possibility of off-target effects and obscured interpretations. Recent technical advances, however, including *in vivo* combined single-unit/LFP recording, chemogenetic, and optogenetic techniques, and activity-based molecular tools, have permitted investigators to begin answering this question. For a full review of the neuronal circuit underlying fear behavior, consult an excellent recent review [47]. In general, it is thought that hippocampus encodes contextual information, mPFC plays a complex role in both forming and

updating associations as well as top-down behavioral control, and the midbrain output centers such as CeA and periaqueductal gray exert control over specific behavioral outputs. A precise delineation of the contribution of each brain region to fear and extinction learning, let alone a description how these contributions arise, remains elusive.

IEGs and Ensemble Theory

While identifying the critical regions in a behavioral circuit is no doubt important, a more fundamental understanding of memory at other levels, in particular the synaptic and cellular levels, is required. The search for physical substrates of memory formation ("engrams") has relied upon and indeed contributed to the prevailing dogma of memory known as ensemble theory [48]. This theory holds that discrete features of the environment, and indeed features of a memory, are encoded in the synchronous firing of discrete sets of pyramidal neurons within a given brain region, although in theory an ensemble can be distributed across multiple regions. A classic example of this are the CA1 "place cells", which fire when an animal is within a certain region or "place" of an environment, and are presumed to encode features of the environment to downstream brain regions when firing together [49]. This theory dates back most notably to the pioneering work of Hubel and Wiesel, who demonstrated that visual cortex neurons respond to discrete visual features of the environment [50]. It is worth pointing out, however, that an ensemble need not merely encode an external feature of sensory input, and indeed it is much more likely that ensembles can be influenced by internal representations. In other words, neurons do not explicitly represent features of the environment. Although the precise definition of a memory encoding ensemble is a matter of some debate, the necessary requirements are

that the subset of cells is active during encoding of a particular memory, and again during the retrieval of that memory, and finally that the firing of the ensemble is sufficient to elicit the downstream effects of feature activation, usually read out via behavioral assays. Recent advances have made direct tests of ensemble theory, long a theoretical construct only, a reality. First, the ability to tag (and later manipulate) neurons that are active during a discrete temporal window has enabled direct visualization of neuronal ensembles [51]. This technique relies on the expression of neuronal immediate early genes (IEGs), a class of proteins which are rapidly expressed after the activation of principal neurons. Multiple IEGs are commonly used as activation markers, and they have slightly different activation requirements and kinetics, but in general they are expressed rapidly in response to robust depolarization of neurons and/or neurotrophic signaling, with protein levels appearing within 30min of activation and peaking between 45-90min [52,53]. The first study to use IEGs to label neuronal ensembles used the IEG c-Fos and studied ensemble encoding within BLA [51]. The authors created a double transgenic mouse, termed "TetTag", which put the tetracycline transactivator downstream of a minimal c-Fos promoter, as well as a LacZ reporter under the control of a tetracycline response element (TRE). This strategy resulted in the labeling of cells expressing c-Fos with LacZ in a doxycycline repressible manner, thereby allowing the experimenter to define a temporal window of ensemble labeling by the absence of Doxycycline. The authors used this transgenic mouse to label neurons active in BLA during fear conditioning, and demonstrated, using another IEG, Zif268, that a significant portion of these cells expressed Zif268, or were "reactivated," during retrieval of the memory. Strikingly, the percentage of reactivation correlated with the fear response of the animal, thereby

providing some of the first strong evidence of sparse ensemble encoding of memory within the BLA, or indeed in any brain region. In another great proof-of-principle leap forward, a more recent study used a similar IEG-based approach to express the light-activated cation channel Channelrhodopsin-2 (Chr2) in cells within the dentate gyrus region of hippocampus which were active during contextual fear learning, and found that subsequent artificial activation of these cells in a neutral context was sufficient to elicit freezing responses [54]. These results, as well as several elegant follow-up studies, have provided the most robust support of ensemble theory, though they are not without experimental caveats [48]. Future studies will need to work toward understanding how the endogenous activity patterns of ensembles represent distinct memory features, and endeavor to mimic physiologic activity when designing optical stimulation protocols. Still, these studies have allowed previously unprecedented access into the mechanisms of memory formation, and have furthered the notion that discrete neuronal ensembles underlie the basis of memory encoding.

Mechanisms of Extinction Learning

Due to their obvious parallels to exposure therapy, the mechanisms of extinction learning have gained much recent interest. Due to that recent interest, a full catalogue of the brain regions involved in extinction learning is now available, and includes amygdala, hippocampus, prefrontal cortex, and other regions known to also be involved in fear conditioning [41,43]. Recent studies have demonstrated that the engagement of parallel inhibitory circuits that could suppress freezing behavior, including the GABAergic intercalated cells (ITCs), and ventromedial prefrontal cortex (vmPFC) [55,56]. A recent

elegant study demonstrated that neurons within BLA that project to vmPFC have opposing activation profiles compared to those that project to dorso-medial prefrontal cortex (dmPFC), a region normally associated with fear behavior [36]. Further, the authors showed that manipulation of BLA→dmPFC or BLA→vmPFC produced opposite effects learning, although they did not look at the acute effect of manipulation. From these and other studies, the predominant notion of extinction learning is that it engages separate, parallel circuits, which are summed in downstream output regions to determine net behavior. In other words, extinction acts as a downstream brake to fear behavior. While simple enough to intuit, this model does not explain how the original association is updated with continued experience, and does not allow for the context and temporal specificity of the learned extinction memory. Furthermore, it does not examine how many previous experiences can be integrated into a single behavioral output. The "on-off" model of fear conditioning and extinction is necessarily limited by these interpretations. Using a version of TetTag, another recent study addressed the mechanisms of extinction learning within the BLA [57]. The authors found that extinction learning reduced the percentage of reactivation among neurons active during fear learning. In addition, they found that, around neurons that were active during fear learning but not active during retrieval (termed "silenced fear neurons"), extinction learning was associated with an increase in the amount of perisomatic parvalbumin (PV) – positive puncta, indicating enhanced inhibitory input from local PV- interneurons. This finding lead to the hypothesis that extinction learning is accomplished, in part, by target-specific changes to inhibitory synaptic weight, resulting in suppression of the fear-encoding BLA ensemble. Interestingly, it also indicates that, within the same local circuitry, discrete and

antagonistic learning events (conditioning and extinction) can occur, leading to a direct interaction of the two memories and continuous updating of associative behavior, a feature likely to be critical to the production of successful behavior strategies. In order to fully understand how changes in perisomatic PV connectivity could produce this feature, a more sophisticated understanding of the role these neurons play in local circuitry is necessary.

PV-Interneurons in Circuit Function

PV-interneurons within the amygdala, and indeed within neocortex and hippocampus, are predominantly fast-spiking GABAergic neurons which innervate the perisomatic region of their targets [58]. This targeting specificity allows GABAergic input from PV-interneurons to prevent or abort excitatory activity propagating from the dendrites, as well as back-propagating calcium waves. They typically express unusually high levels of NaV channels in their axonal compartment, and are morphologically structured to produce highly faithful action potential firing [59]. In addition, a single PV-interneuron in BLA innervates between 800-1000 principal neurons [60]. Due to these features, PV-interneurons are designed to exert extremely tight inhibitory control many targets, and are thus positioned to exert both global gain modulation and specific ensemble-grouping functions.

For further insight into the role of PV-interneurons in BLA, a brief review of canonical inhibitory circuit motifs is useful. Most commonly, inhibitory circuit motifs can be classified as recurrent (or feedback), lateral, or feedforward. Recurrent inhibition is characterized by an excitatory and inhibitory neuron pair forming a microcircuit, in

which the excitatory neuron directly excites the interneuron, which in turn inhibits the starting excitatory neuron. Therefore, in simplest terms, principal neurons generates its own inhibition with a disynaptic delay through an interneuron intermediate. A related form of inhibition called lateral inhibition is characterized by a principal cell or ensemble driving the inhibition of a separate cell or ensemble through an interneuron intermediate. This motif is generally thought to be the fundamental basis for ensemble segregation, as it allows discrete sets of principal neurons to be active while silencing adjacent sets [58]. Finally, in the third form of inhibition, feed-forward, interneurons are driven not by local sources, but by long-range excitatory input, thereby allowing distant brain regions to affect local computations. In truth, in cortical and cortical-like brain regions, all three of these motifs are omnipresent, and in fact often converge on a single interneuron hub. A good example of this is, in fact, PV-interneurons in BLA, which predominantly receive their excitatory input from local principal neurons, but also participate in feedforward inhibition as well as disinhibitory circuits that are critical for learning [29,61]. Therefore, the relative significance of these motifs and their ultimate effects on circuit output is determined by a variety of parameters, most important of which are the relative weights of principal neuron to interneuron and interneuron to principal neuron synapses – a parameter modified by extinction learning in the BLA – as well as, critically, the timing of the inputs to the interneuron hubs relative to the excitatory inputs to their targets. The fact that timing of inputs is paramount is obvious, yet often missed in static visualization of circuit properties. This point is nicely illustrated by recent work addressing the effects of inhibition on principal neurons within BLA, in which the authors found prominent rebound spiking – that is, spiking of principal neurons following hyperpolarization – in

BLA after compound IPSPs from perisomatic-targeting PV interneurons [62,63]. Interestingly, this phenomenon was dependent on coordination between sub-threshold membrane oscillations and IPSPs, thereby limiting the firing window of principal neurons to a dedicated frequency band. It is not hard to imagine how such a circuit property could lead to the coordinated activity of subsets of principal cells in a given frequency band, and how learning could alter ensemble firing by modifying parameters of this integration. Recent work *in vivo* has demonstrated similar phenomena in cortex, and it is increasing clear that inhibition plays a critical role in the coordination and synchronization of principal neuron ensembles [64,65].

Oscillations and Circuit Function

An outstanding question from the previous section is how the coordination of neuronal synchrony is regulated on timescales necessary to support cognitive functions. While biochemical changes in synaptic weight, such as classical long-term potentiation (LTP), would be predicted to alter the circuit motifs listed above, other parameters must exist to allow the rapid grouping and mixing of neuronal ensembles during behavior, as well as the effective and flexible communication with distant circuits. A proposed solution to these problems lies in the emergent circuit property of oscillations. Defined minimally, an oscillator is a physical system that exhibits periodicity. The biochemical network that supports circadian rhythms is a classic example. From the first time electrical signals were recorded from human and animal brains, it was clear the large scale brain activity was represented in oscillations, and electroencephalographic oscillatory features have long been used to identify and study disorders of human neural

circuit function such as epilepsy [66,67]. Whether recorded by EEG or by depth electrodes, local field potentials (LFPs) are the vector sum of all potentials within a given field, weighted by proximity to the recording electrode. As such, they represent the sum of activity in a given brain region or local circuit, rather than the activity of a specific cell, as is the case in single-unit recordings, and provide a window into the circuit level (as opposed to synaptic or cellular *per* se) operations of the brain [68]. The fact that oscillations are salient features of brain activity across many different brain regions and indeed many species of animal likely means that oscillatory activity plays a critical role in supporting complex brain function. Indeed, the different frequencies of oscillations have long been correlated with particular behaviors or cognitive tasks, and one of the common tasks of early systems neuroscience was to catalogue each oscillation based on frequency, brain region, and associated behavior, as to infer function from them [69]. As it turns out, the outward-in approach of starting from the behavior in order to deduce the function of emergent neural phenomena is a near impossible task, not just for oscillatory phenomena. The more fundamental and pressing question, however, is *what* exactly do oscillations *do?* Then we can begin to apply that function to particular output behaviors or sensory inputs in a systematic manner. Thanks to pioneering work from several groups, as well new technologies capable of more sophisticated correlative studies, as well as gain-of-function studies, this question is beginning to be answered.

The question "how do oscillations arise?" is far beyond this scope of this modest review, but suffice it to say that they are a prototypical complex emergent phenomenon; they critically rely on, and are modified by, the interaction between the lower level components of the circuit such as synaptic weight and delay, input timing, circuit motifs,

neuromodulation, etc [69]. In addition, it is worth pointing out that oscillatory activity is impossible without inhibition, which provides the countervailing force to feedforward excitatory activity and endows neuronal circuits with their characteristic non-linearities [65,70]. In fact, the lone criterion for an oscillation is simply the tug-of-war between excitatory and inhibitory drive, with some offset time delays. A simple example would be a feedback inhibitory loop, as described above, where cell A drives firing of cell B, which in turn inhibits cell A. Application of noisy input into the system from an external source would lead to an periodic pattern of action potential firing, with its frequency being dependent on the synaptic and refractory delays between the two components. A balance between excitatory and inhibitory processes, as seen here, is thought to underlie all neural oscillations. It is also worth noting, too, that oscillations are not merely emergent from the substituents of the circuit. Rather, they emerge from and subsequently constrain the activity of those same substituents, making them a critical regulatory hub in circuit activity and a nexus of hard to predict, non-linear relationships[69]. Just as one cannot reverse engineer complex brain behavioral output by focusing purely on substituent levels (eg: synaptic function) and ignoring the properties of the behavior *per se*, one cannot understand circuit dynamics without addressing the phenomena of the oscillations themselves[69].

Brain oscillations are generally classified as relaxation-type, meaning that they have distinct accrual (integration) and duty (output or discharge) phases. A commonly used and intuitive real-world analogy of this is a leaky faucet, which slowly accrues enough water until the droplet is sufficiently large to cause it to fall (discharge) due to gravity, thereby restarting the cycle[69]. Though this behavior is periodic, the accrual and

duty phases need not be equivalently long in duration. Also, relaxation oscillators, like the faucet, exhibit the critical property of resetting. An external force, like a finger tapping on the faucet, or external input from another brain region, can cause premature discharge, thereby resetting the oscillation so that that in may now have a different phase relationship with other oscillators. For this reason, coupled relaxation oscillators tend to synchronize with each other naturally, thereby providing a means of efficient information transfer. Critically, this synchronization can be toggled by other external inputs, allowing for rapid synchronization and desynchronization of local and distant circuits, essentially providing a means for selective "listening of inputs" [71,72]. Already from this simple (and insufficient) description we can see that two of the main functions of neural oscillators arise. First, due to segregation of accrual and duty phases, oscillations are a means by which circuits can sort groups of cells into ensembles by linking them to a particular ongoing oscillation and therefore to that oscillation's phase relationships. That this may seem similar to the function of feedback inhibition stated above is not coincidental; indeed interneurons are thought to underlie most oscillatory activity in the brain, including gamma oscillations known to be essential for ensemble segregation [65,73]. Second, because of this separation of accrual and duty phases, circuit output is naturally parsed into interpretable bits of information for downstream circuits. This feature, coupled to the means of toggling between synchronization states, means that oscillations are the primary means of regulating interregional communication. These two points are illustrated nicely in recent enlightening work from the Herry lab, among others. In two studies, their group identified an oscillation in the 4Hz frequency range that synchronizes prefrontal cortex and amygdala during fear behavior, is generated by synchronous PV-

interneuron activity, and is causally related to the behavior [74]. They then demonstrated, through a combination of LFP and single-unit recordings, that cell ensembles in mPFC that correlate with fear behavior are grouped in the ascending phase of the oscillation, and optogenetic manipulation of this phase relationship can toggle the downstream behavioral effect of the oscillation [75]. Though the literature on this general subject is vast, in particular regarding theta oscillations in the hippocampus, this study remains a particularly strikingly example of the fundamental synchronization and ensemble grouping functions of oscillatory activity.

Oscillations and Fear Learning

From the work of the Gordon, Pape, and other groups, it is clear that both fear learning and extinction modify the oscillatory parameters of the hippocampo-cortico-amygdalar limbic system. An incomplete list of these parameters includes theta (4-10Hz) band synchronization of hippocampus and amygdala during retrieval of a conditioned fear memory, alterations of the directional communication in theta frequencies among the limbic structures following both fear conditioning and extinction learning, and alterations in both fast-gamma (70-120Hz) power and phase-amplitude coupling to theta frequencies within amygdala following extinction learning [76–80]. From this important work, we can clearly deduce that oscillatory phenomena play critical roles in sculpting fear circuit activity, and are modified by learning experiences. We cannot, however, say exactly what these oscillatory parameters do without a more sophisticated approach. Namely, with purely correlative approaches such as those listed, we are limited in the scope of our interpretations. Only the aforementioned work from the Herry lab has heretofore

convincingly demonstrated a causal role for the oscillation itself, by controlling for overall levels of activity and other frequency-band manipulations. Still though, the outstanding question from all of this work remains what function the oscillation *per se* provides, and how it accomplishes this. To address these fundamental questions, interrogation of the circuit components that give rise to the oscillation, their modification through learning, and the top-down effect the oscillation has, in turn, on these components, is absolutely critical. Through the combined use of chemogenetic/optogenetics, functional ensemble labeling, slice-electrophysiology, *in vivo* LFP recording, circuit tracing, and behavior analyses, I have attempted to accomplish a small part of just that. By observing how a circuit undergoes learning-dependent changes across multiple levels, and manipulating critical circuit hubs both before and after learning, I have aimed to link the somewhat disparate fields of ensemble memory encoding and oscillatory circuit phenomena, and address the simple question: how is extinction learning encoded in the cortico-amygdala circuit?

Chapter 2: Materials and Methods

Animals

All animal procedures were performed in accordance with the NIH Health Guide for the Care and Use of Laboratory Animals and were approved by the Tufts University Institutional Animal Care and Use Committee. The TetTag mice used in this study were heterozygous for two transgenes: c-fos promoter- driven tetracycline transactivator (cfosP-tTa) and a tet operator-driven fusion of histone2B and eGFP (tetO- His2BGFP). PV-Cre and PV-Cre/TetTag mice used in this study were heterozygous for a PV-IRES-Cre knock-in locus (B6; 129p2-Pvalbtm1(cre)Arbr/J). Mice had food and water ad libitum and were socially housed until the start of behavioral experiments, which was at an age of at least 12 weeks. Mice were kept on regular light-dark cycle, and all experiments were performed during the light phase.

Stereotaxic surgery

Mice were anesthetized with isoflurane, held in a stereotaxic apparatus (Kopf), and injected with virus. After injection, the needle was left in place for 10 min before slowly retracting. The incision was sutured, and mice were weighed and monitored to ensure recovery. For DREADD experiments, we injected 400nl of AAV-Syn-DIO-hm4Di-mCherry (UNC Vector Core, Bryan Roth) bilaterally into basal amygdala (AP −1.35 mm, ML ± 3.45 mm, DV −5.15 mm). For anterograde tracing of BLA!mPFC projections, 200nl of AAV9-CaMKII-Cre-GFP was injected into mPFC (AP +0.75, ML ± 0.3mm, DV -2.1 mm) and 400nl of AAV-Syn-DIO-hm4Di-mCherry was injected into BLA.

For rabies tracing in PV-Cre mice, 250nl of AAV-EF1a-FLEX-GTB "Helper virus" was injected to basal amygdala, and glycoprotein deleted rabies virus (SADΔG_mCherry(EnvA), generously provided by Dr. Ed Callaway and the Salk Institute Vector Core, was mixed with CTB647 for injection site localization and injected 3 weeks later into the same region. Mice were sacrificed and tissue was collected and processed 5- 9 days following rabies injection. All coordinates are relative to bregma.

Behavior

3-6 month old PV-Cre and PV-Cre TetTag mice were used for the study. For DREADD experiments, each group consisted of a group of mice receiving CNO injections followed by Vehicle (CNO/VEH), and another group with the reverse order (VEH/CNO).

The design of experiment 1 is summarized in Figure 1. Mice were subjected to contextual fear conditioning consisting of three training trials (S1, S2, and S3) with 3 hours between each trial. The total duration of each training trial was 500 s. A training trial started with placing the mouse in a square chamber with grid floor (context A) (Coulbourn Instruments; H10-11RTC, 120W 3 100D 3 120H). At 198 s, 278 s, 358 s, and 438 s, a foot shock was delivered (2 s, 0.70 mA). On days 2 and 3 (or 4 and 5 for EXT Tagging group), mice were subjected to four extinction trials per day. Each extinction trial lasted 1200s with a trial interval of 2 hr. For each extinction trial, mice were placed in the same box used for fear conditioning without receiving foot shocks. On days 4 and 5, mice were tested over 500 s during a single retrieval test in context A. A subset of mice

in Experiment 1 also were tested in a neutral context (Context B), which consisted of a square plastic box with bedding sprayed with 10% acetic acid and striped walls. CNO/VEH mice received an IP injection of 8-10mg/kg CNO on day 4 and of Vehicle (5% DMSO in saline) on day 5. VEH/CNO received injections in the reverse order. All TetTag mice were perfused 90 min after the final retrieval trial (day 5 for FC tagging, day 7 for EXT tagging).

For labeling of activated cells, TetTag mice were raised on food with doxycycline (40 mg doxycycline/kg chow). For "FC Tagging" group, 4 days before conditioning, mice were individually housed, and doxycycline was removed from the food. After the last fear conditioning trial on day 1, mice were put on food with a high dose of doxycycline (1 g/kg) to rapidly block the tagging of neurons activated after fear conditioning. On day 2 mice were put back on the regular dose of doxycycline (40 mg/kg). For "EXT Tagging", mice were conditioned while on doxycycline in an identical fashion to "FC tagging" group. Doxycycline was subsequently removed from food and animals remained in homecage for 2 days before beginning two days of extinction trials. Following the last extinction trial on day 5, mice were put on with high doxycycline (1g/kg).

For no-shock vs. shock comparison, a separate cohort of mice was split into 3 groups. Homecage mice were taken off dox for four days and and left in homecage. Context + shock (FC) mice were taken off dox for four days prior to contextual fear conditioning as described above, then placed back on high doxycycline chow as described above. Context no-shock mice were placed in the conditioning context for three

identical trials with the exception that no shock was given. No-shock and FC mice were perfused 90min following a retrieval trial the next day.

Quantification of Freezing

Freezing behavior was measured using a digital camera connected to a computer with Actimetrics FreezeFrame software. The bout length was 1.5 s and the threshold for freezing behavior was set prior to the onset of the experiment, and was the same for all subsequent trials for each animal. Freezing scores were obtained by averaging freezing during the entire trial, unless otherwise indicated.

Freezing for optogenetic experiments was scored manually due to light interference with video monitoring. Scoring was performed by lab member blinded to the experimental design. “No Stim” freezing data included seconds 0-40 and 210-240. 4Hz Stim and 8Hz stim freezing scores were averaged over 2 30sec stim sessions per trial, unless otherwise indicated. Same for LFP analysis unless otherwise indicated for artifact removal. Mice were included in this if 1) proper virus expression in BLA 2) sufficient extinction learning as measured by Ret/EXT1 ratio < 0.4 and 3) optic cannula was located close to BLA, as determined by proximity to recording electrode post-hoc.

Tissue Preparation and Immunohistochemistry

Ninety minutes after retrieval, mice were deeply anesthetized with ketamine/xylazine and intracardially perfused with 0.1 M phosphate buffer (PB) followed by 4% paraformaldehyde (PFA 4%) dissolved in 0.1 M PB. Brains were extracted and post-fixed in PFA 4% for 24 hr. Brains were transferred to 30% sucrose for 48–72 hours before slicing 40mm (for hM4Di-mCherry localization and Nissl staining) or 25 mm (for all immunofluorescence) coronal sections of the entire brain using a cryostat. Sections were stored in PBS with 0.025% sodium-azide at 4°C until use. For immunoflouroscent staining, sections were blocked for 1 hr at room temperature in PBS-T with 8% normal goat serum. Sections were incubated in rabbit anti-Zif268 (Santa-Cruz; polyclonal; 1:3,000) with mouse anti-PV (Millipore; monoclonal; 1:2,000), or rabbit anti-RFP (Rockland; 1:1500) at 4°C for 48-72 hours. Secondary antibodies (Jackson ImmunoResearch; goat anti-rabbit 549 1/1,500, goat anti-mouse 647 1/500) were diluted in the blocking solution and were then applied to the sections for 2 hours at room temperature followed by three rinses for 15 min in PBS-T. Sections were mounted on slides and coverslipped after a brief wash with 0.00005% DAPI in PBST to label cell nuclei and stored at 4°C.

Microscopy

A wide-field epifluorescent microscope (Keyence BZ-X700) was used to acquire images for electrode and injection site validation, and for TetTag data excluding perisomatic mCherry analysis. 10-20x images were obtained and stitched together using Keyence software. Acquisition settings were optimized for each brain region and were

identical across groups. A minimum of 8 total amygdalae or 6 other brain region sections (4 or 3 bilateral) per animal was analyzed, after excluding sections for quality reasons. Sections including damage from injection site were excluded.

A confocal laser-scanning microscope (Nikon A1R, or Leica SPE) was used to acquire images for mCherry/PV overlap analysis and mCherry-Zif correlation analysis. The settings for PMT, laser power, gain, and offset were identical between experimental groups. For mCherry/PV overlap, 20x z-stacks were acquired and the maximum intensity projection was used for analysis. For mCherry-Zif correlation analysis, 20x images were used. A minimum of 4 sections of amygdala per animal was used for this analysis.

Quantification of Activated Cells

ImageJ software was used to select and count the total number of DAPI-, GFP-, and Zif-positive nuclei and nuclei double positive for GFP and Zif . In order to avoid bias, all three cell types (GFP+Zif−, GFP+Zif+, GFP−Zif+) were selected from the same pictures, and the threshold settings for GFP and Zif were identical across all mice.

Mice were included in this analysis based on the following criteria: 1) Behavioral: average freezing in EXT8 trial must be <50% of freezing in EXT1 trial 2) TetTag: There must be at least an average of 10 GFP+ cells in BLA per section.

Perisomatic analysis

Selection of GFP-labeled cells for confocal perisomatic analysis was designed to only include excitatory neurons as described previously[7]. mCherry signal was distinguished from nuclear ZIF using ImageJ image calculator function and DAPI as a mask for nuclear signal. Absolute values of fluorescence were calculated for each individual neuron in an unbiased way without knowledge of GFP+ fluorescence or experimental group. Values were then normalized within each section to reduce noise from imaging and tissue processing and allow for comparison across animals.

Electrophysiology

All LFP data were acquired using pre-fabricated headmounts (Pinnacle 8201) for 2EEG/1EMG recordings. The headmounts are affixed to the skull with stainless steel screws which also act as EEG reference and ground electrodes. Stainless steel wires serve as the EEG electrodes, which are placed into the BLA and mPFC (BLA AP −1.35 mm, ML ± 3.45 mm, DV −5.15 mm, mPFC AP +0.75, ML ± 0.3mm, DV -2.1 mm). Local field potentials were recorded using a 100x preamplifier (Pinnacle 8202-SE) at a sampling rate of 4kHz. The data were analyzed and filtered offline using LabChart software. Electrode placement was assessed by histological analysis of nissl-stained 40 micron sections and only mice with BLA placement were used for further analysis.

A total of 12 mice were included in LFP analysis. Mice were excluded from various analyses based on analysis-specific criteria. Of these 12, only 9 had correct mPFC

placement as well. Mice were excluded from fine temporal analysis (see below) if we did not have temporally matched freezing files for that trial.

All power spectra quantification data were generated using LabChart's DataPad function, using the first 4 minutes of each behavioral trial. Normalized power and 3-6:6-12Hz ratio (Figure 3E) were calculated by subtracting the power or power ratio value for the first 2 minutes of the first conditioning trial, during which the animal is acclimating to the context, from the ratio for the CNO or VEH retrieval trials.

Directionality was calculated using a modified process as described by others. In short, instantaneous amplitudes of filtered traces from mPFC and BLA were generated in LabChart. Files were then exported into ClampFit and cross-correlated after being segmented into 20s bins, and statistics were calculated using the population distribution of lags per individual trial (based on the first 4 minutes of each trial). For finer temporal scale directionality analysis, periods of at least 10s of motion followed by at least 10 seconds of freezing (one per animal per trial) were selected to maximize signal:noise. The onset of freezing was demarked 0s and directionality analysis was performed from -5 to 5s in 1s bins with 0.5s steps. Lags were then averaged across animals on a per-second basis for visualization, or, for quantification, for the 5s prior to freezing and the 5s after the onset of freezing. For pure freezing analysis, separate files were created in LabChart collating all periods of pure freezing or non-freezing for a given trial using second-by-second freezing data from FreezeView. These files were then subsequently used for analysis as indicated. Mice were excluded from this analysis if, during retrieval trials, they did not exhibit freezing behavior sufficient (as described above) for inclusion.

Sub-band cross-correlation analysis was performed in ClampFit by cross-correlating instantaneous amplitudes of 3-6Hz and 6-12Hz bands over a 2 min window at beginning of trial and per each trial per animal. Statistics were performed using area under curve of cross-correlogram. Onset of freezing analysis was performed using 1s bins from the same freezing onset epochs as described above. Spectrograms and coherence analysis were generated using the EEGLAB plugin in MATLAB [81]. Quantification of peak coherence for DREADD experiments was performed using the first 4 minutes of each trial, and exporting coherence curves to GraphPad Prism for analysis. Coherence for optogenetic experiments was calculated using the first 40seconds (No stim) or pooling periods of stimulation (4Hz and 8Hz stim).

In-vivo optogenetic manipulation of PV-interneurons

1-2 weeks after unilateral BLA infusion of AAV-DIO-Chr2-mCherry, optic cannulae (THOR Labs, part# CFM22L05) were targeted to BLA using a stereotaxic setup similar to that used virus infusion. Cannulae were affixed to the skull with dental cement during electrode/headmount implantation procedure. Fiberoptic cables and pre-amplifiers were attached to animals one day prior to the behavioral paradigm in order to give the animals time to acclimate, and were kept attached for the duration of the experiment as to avoid behavioral confounds due to excessive handling of the animals. The stimulation protocol for all trials was as follows: after 40s of no stimulation, 4 alternating 30s periods of 4Hz or 8Hz sinusoidal stimulation (10mV peak amplitude) with 20s of no stimulation in between. The order of 4 or 8Hz stimulations was reversed in approximately half of the animals, but was consistent across all trials for a given animal.

Slice Electrophysiology

PV-Cre:TetTag mice were infused bilaterally with 400nl of AAV-DIO-Chr2-mCherry into the BLA as previously described. Mice were put through "FC-Tagging" and "EXT-Tagging" paradigms as previously described. 3-7 days after behavior, brains were rapidly removed following decapitation under isoflurane anesthesia and immersed in ice cold artificial cerebral spinal fluid (aCSF) containing (in mM) 126 NaCl, 26 $NaHCO_3$, 1.5 NaH_2PO_4, 5 KCl, 2 $CaCl_2$, 10 dextrose (300–310 mOsm) and 3 mM kynurenic acid. Coronal sections (350 μm) were cut using a Leica vibratome and incubated at 33°C in normal aCSF for at least 1 h prior to transferring to the recording chamber. The recording chamber was maintained at 33°C (in line heater, Warner Instruments) and continuously perfused at a rate of ≥4 ml/minute with aCSF throughout the experiment. Solutions were continuously bubbled with 95% O_2 and 5% CO_2.

Patch clamp measurements from visually identified TetTag or dark BLA pyramidal neurons were made using a 200B Axopatch amplifier (Molecular Devices) and recorded using PowerLab Hardware and LabChart 7 data acquisition software (AD Instruments). Borosilicate glass micropipettes (World Precision Instruments) with DC resistance of 5–8 MΩ were backfilled with internal solution containing (in mM): 140 KCl, 0.5 EGTA, 5 HEPES and 3 Mg-ATP. Adjust pH to 7.3 with KOH. Adjust osmolarity to 300 mosM with glucose.

Intrinsic resonance properties were determined in current-clamp mode by application of a sub-threshold (appx 50pA) sinusoidal current of linearly increasing frequency and

constant amplitude. Voltage responses were measured, pooled from multiple recordings from a single cell, and the resonance properties were determined by calculating the FFT of the voltage response (assuming constant input amplitude).

Firing rate measurement were made by injecting a step current to depolarize the neuron to threshold. Cells were held at threshold and average firing rate was determined for baseline (no stimulation), and during 4Hz or 8Hz sinusoidal optical stimulation of PV-interneurons. Normalized firing rates were calculated using the ratio of the firing rate during stimulation to the ratio of the 30s prior to stimulation, in order to partially account for cell fatigue and other possible sources of altered neuronal behavior following multiple manipulations.

Statistics

Statistical tests were performed using Prism (GraphPad) and are indicated in the figure legends. All statistical tests were two-tailed. Assumptions about normal distributions were tested using the Shapiro-Wilk normality test, and assumptions about equality of variances were tested using an F-test. Non-parametric tests were used when assumptions for the parametric test were not met (unpaired t-test: normal distribution and equal variance of population values; paired t-test: normal distribution of differences; linear regression: normal distribution of residuals). Box plot graphs show median, 25% and 75% percentiles as box edges, and minimum and maximum values as error bars.

Chapter 3: Results

Section 1: PV-interneurons underlie critical features of extinction learning with the BLA

Because previous work in our lab had associated extinction learning with target-specific structural changes in PV+ puncta surrounding BLA "FC neurons" – that is, neurons active during fear conditioning within the BLA -- we wanted to directly test the role of PV-interneurons in the suppression of fear behavior following extinction learning. In order to do so, we injected AAV-Syn-DIO-hM4Di-mCherry, an AAV which expresses the inhibitory Designer Receptor Exclusively Activated by Designer Drug (DREADD) receptor hM4Di in a Cre-dependent fashion, into the BLA of PV-IRES-Cre mice, which express Cre recombinase in PV-interneurons. hM4Di is a modified Gi metabotropic receptor that, when bound by its exogenous ligand clozapine-N-oxide (CNO), decreases adenylyl cyclase activity and activates inward rectifying potassium channels, thereby resulting in hyperpolarization and a reduction presynaptic release probability [82]. Using this strategy, we were able to express hM4Di selectively in BLA PV-interneurons, and exert functional control over their firing with spatial and temporal specificity (Figure 1.1 A, Figure 1.2 A-B).

After allowing time for sufficient expression of the virus, we subjected animals to our contextual fear conditioning and extinction paradigm (Figure 1.1 B), and tested the effect of silencing BLA PV-interneurons during a subsequent retrieval trial. We found

that silencing BLA PV-interneurons during retrieval of an extinction memory resulted in a increase in freezing responses in the conditioned context, but not in a neutral, unconditioned context, implicating PV-interneurons in the context specific suppression of freezing following extinction learning (Figure 1.1 D). Given our previous finding that extinction learning seems to elicit target-specific structural plasticity of the BLA PV-interneuron network, we reasoned that perhaps extinction learning itself was necessary to confer PV-interneurons with their dedicated role in context-specific fear suppression. To test this, we used the same DREADD-based approach as above, while subjecting animals to a modified behavioral paradigm in which they were fear conditioned but not put through extinction learning. Strikingly, silencing BLA PV-interneurons during retrieval had no effect on freezing levels, indicating that the BLA PV-network plays no role in the modulation or suppression of fear behavior before extinction learning (Figure 1.1 E).

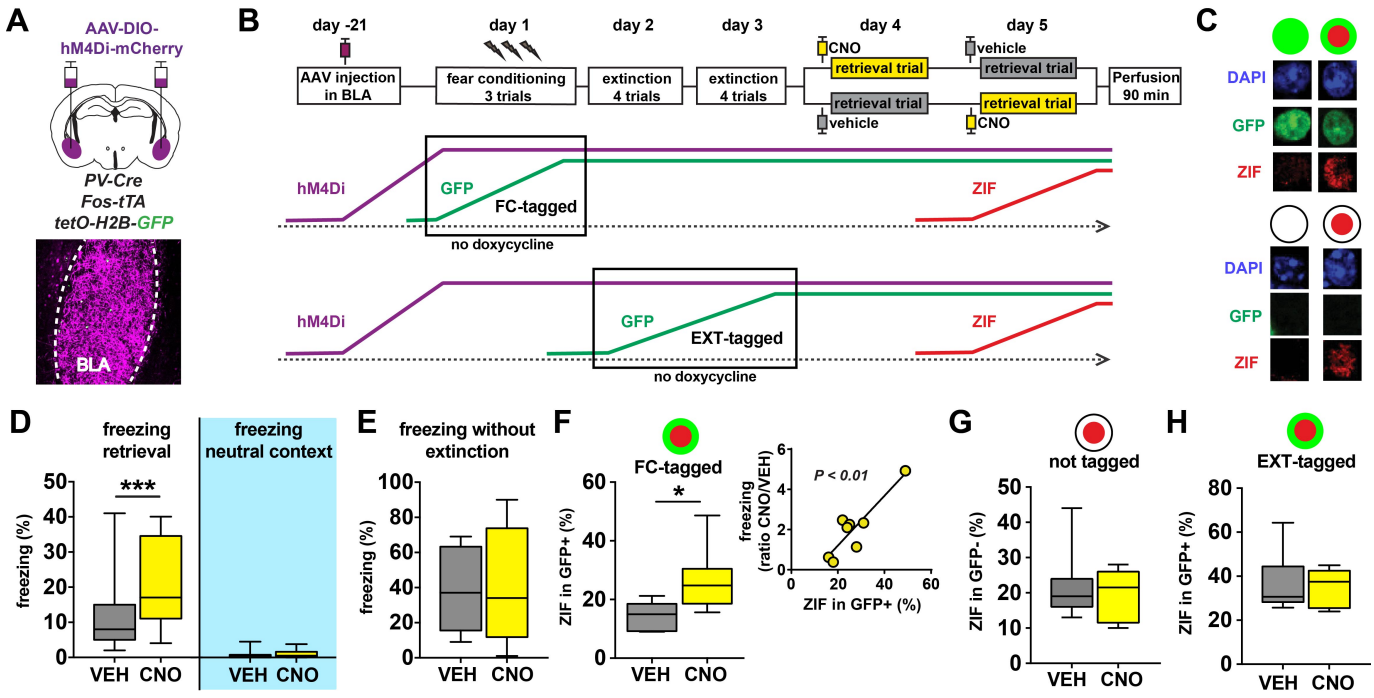

Figure 1.1: BLA PV-interneurons selectively suppress conditioned fear behavior and neuronal ensembles following extinction

A) Bilateral infusion of AAV-Syn-DIO-hM4Di-mCherry into the BLA of PV-Cre:Fos-tTa:tetO-Hist1H2B-GFP (PV-Cre:TetTag) mice was used to selectively express DREADD receptors in BLA PV-interneurons. **B)** Mice were subjected to contextual fear conditioning, extinction, and retrieval. Cells active during the no doxycycline period are tagged with long-lasting H2B-GFP expression as a result of Fos-promoter driven tTA expression, which enables tTA protein to bind and activate the tetO-H2B-GFP transgene. This binding is prevented in the presence of doxycycline. Cells active during the retrieval trial express short-lasting Zif268 (ZIF) protein that can be detected in brains perfused 90 minutes later. PV-Cre:TetTag mice were without doxycycline chow during either the fear conditioning trials or during the extinction trials, resulting in FC-tagged and EXT-tagged neurons, respectively. For the EXT-tagged experiment, extinction and retrieval was done on days 4-7 instead of days 2-5 to enable clearance of doxycycline on days 3-4. **C)** Examples of GFP+/GFP- and ZIF+/ZIF-nuclei in the BLA of TetTag mice. **D-E)** Silencing BLA PV-interneurons results in a selective increase in conditioned fear following extinction. Mice injected with CNO 30 min prior to retrieval displayed increased freezing in the conditioned context (**D left**, Wilcoxon matched-pairs: $W = 338$, $P < 0.0001$, $n = 29$ mice), but not in a neutral context that was never paired with footshock (**D right**, Wilcoxon matched-pairs: $W = 9$, $P = 0.6563$, $n = 9$ mice). Mice injected with CNO 30 min prior to a retrieval trial that was not preceded by extinction trials did not show altered freezing levels in the conditioned context (**E**, Wilcoxon matched-pairs: $W = -1$, $P = 0.9824$, $n = 12$ mice). **F-H)** Silencing PV interneurons selectively disinhibits the BLA fear ensemble. Injection of CNO 30 min prior to retrieval leads to an increase in the percentage of ZIF+ cells among the FC-tagged GFP+ neurons (**F left panel**, unpaired t-test: $t(13) = 2.86$, $P = 0.0134$, VEH $n = 7$ mice, CNO $n = 8$ mice). The percentage reactivated GFP+ cells correlated with the behavioral effect of CNO (**F right panel**, linear regression: $F(1,6) = 23.42$, $P = 0.0029$, $n = 8$ mice). Injection of CNO had no effect on GFP- neurons that were not tagged during fear conditioning (**G**, Mann-Whitney test: $U = 27$, $P = 0.9333$, VEH $n = 7$ mice, CNO $n = 8$ mice), or on EXT-tagged GFP+ neurons (**H**, unpaired t-test: $t(10) = 0.4142$, $P = 0.6875$, VEH $n = 7$, CNO $n = 5$).

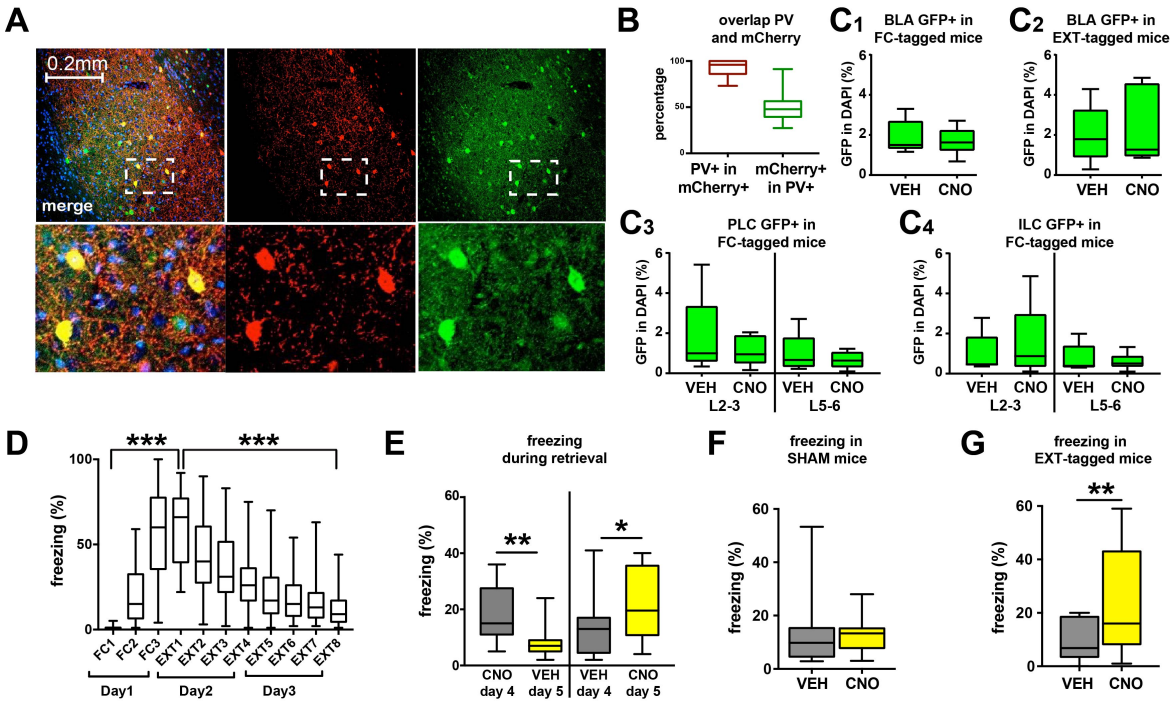

Figure 1.2: Selective expression of hM4Di-mCherry in BLA PV+ interneurons mediates behavioral effect. **A)** Representative 10x confocal z-stack of BLA showing selective expression of hM4Di-mCherry in parvalbumin-positive (PV+) cells (red: mCherry, green: parvalbumin). **B)** Quantification of % overlap between hM4Di-mCHerry and PV (n = 10). **C)** The number of tagged GFP+ neurons did not differ between VEH and CNO injected mice (**C1**, unpaired t-test: t(13) = 0.7906, P = 0.4433, VEH n = 7 mice, CNO n = 8 mice; **C2**, unpaired t-test: t(10) = 0.2709, P = 0.7920, VEH n = 7 mice, CNO n = 5 mice; **C3 L2-3**, Mann-Whitney test: U = 21, P = 0.8981, VEH n = 5 mice, CNO n = 9 mice; **C3 L5-6**, Mann-Whitney test: U = 21, P = 0.8981, VEH n = 5 mice, CNO n = 9 mice; **C4 L2-3**, Mann-Whitney test: U = 20, P = 0.7972, VEH n = 5 mice, CNO n = 9 mice; **C4 L5-6**, Mann-Whitney test: U = 22, P > 0.9999, VEH n = 5 mice, CNO n = 9 mice;). **D)** Mice exhibit increasing freezing levels as conditioning progresses on day 1, and decreasing freezing levels as extinction progresses on days 2-3 (Wilcoxon matched-pairs FC1 versus EXT1: W = 435, P < 0.0001, n = 29 mice; Wilcoxon matched-pairs EXT1 versus EXT8: W = -435, P < 0.0001, n = 29 mice). **E)** Mice freeze more during retrieval after CNO injection regardless of order of trials (CNO first, Wilcoxon matched-pairs: W = -75, P = 0.0015, n = 13 mice; VEH first, Wilcoxon matched-pairs: W = 77, P = 0.0264, n = 16 mice). **F)** CNO has no effect on behavior in animals not expressing hM4Di (SHAM injection) (Wilcoxon matched-pairs: W = 7, P = 0.7344, n = 9 mice). **G)** Mice exhibit increased freezing levels after CNO injection compared to VEH injection in the EXT-tagged group (Wilcoxon matched-pairs: W = 75, P = 0.0059, n = 13 mice).

Because this result runs counter to the simplistic notion of interneurons as "inhibitory brakes" on excitatory behavior, we decided to further address the mechanistic underpinnings of PV-interneuron function following extinction learning. In order to do this, we combined the above chemogenetic approach with the functional ensemble labeling in TetTag transgenic mice, such that we could address the relationship between PV-interneuron activity and the activation state of functional ensembles within the BLA. By labeling FC neurons during conditioning, and then measuring the expression of the neuronal IEG Zif268 ("ZIF") following a retrieval trial, we were able to track the activation state of discrete functional ensembles across multiple timepoints (Figure 1.1 C-D) [51]. We found that the silencing of PV-interneurons during extinction retrieval lead to an increase in the percentage of GFP+ FC neurons that expressed ZIF within the BLA, indicating increased number of "reactivated" FC neurons (Figure 1F). Interestingly, we observed no such effect in GFP-, non-ensemble neurons, indicating that PV-interneurons play a critical role in the selective suppression of FC neurons following extinction learning (Figure 1.1 G). In addition, we found that the percentage of reactivation of FC neurons correlated with the behavioral effect of PV-interneuron silencing across animals, supporting the notion that PV-interneurons reduce freezing levels following extinction in part by suppressing the activity of the original "fear memory encoding" neurons within the BLA. An alternative explanation would be that baseline differences in intrinsic excitability bias a subset of BLA principal neurons to express GFP, and this same subset would therefore be more likely to be reactivated following PV-interneuron silencing. To

test against this, we performed a separate experiment in which we labeled neurons active during extinction learning ("EXT neurons"), and performed the same PV-interneuron manipulation during retrieval trials. We found that, although these mice froze more following PV-interneuron silencing, there was no difference in reactivation percentage of EXT neurons, indicating that PV-interneurons do not play a critical role in regulating the activation state of this distinct ensemble (Figure 1.1 H, Figure 1.2 G). Taken together, these data indicate that extinction learning confers the BLA PV-interneuron network with a role in context-specific fear suppression, which is in part accomplished through the selective control of functional ensembles within the BLA.

To further address how PV-interneurons are capable of selective ensemble suppression, we quantified the relationship between perisomatic mCherry puncta, which correlated tightly with PV-puncta, and ZIF expression within individual neurons (Figure 1.3 A-B). We found that, in vehicle-injected animals, there was no relationship between perisomatic input and ZIF expression in FC neurons, consistent with a role for PV-interneurons in normalizing overall activation within the BLA. We found, however, that silencing PV-interneurons lead to a positive correlation between perisomatic puncta and ZIF, indicating that the neurons receiving the most input from PV-interneurons were the most likely to be reactivated following chemogenetic silencing (Figure 1.3 C). This somewhat counterintuitive finding can be interpreted as the result of disinhibition; the neurons receiving the most perisomatic innervation are the most likely to be disinhibited. Strikingly, however, this effect of PV-interneuron silencing was specific to FC neurons, as we observed no such shift towards a positive correlation in either GFP- or EXT neurons (Figure 1.3 D-E). We reasoned that this discrepancy could be accounted for by

differences in perisomatic innervation among the functional BLA ensembles, but found that FC neurons and EXT neurons received broadly the same amount of perisomatic innervation (data not shown). In other words, while the amount of synaptic puncta, a proxy for synaptic input, explains the effect of PV-interneuron silencing on FC neurons, and confirms that our manipulation indeed acts through disinhibition of perisomatic-targeted FC neurons, it does not explain the selective disinhibition of FC neurons compared to other ensembles within the BLA. One possible explanation for this discrepancy is FC neurons simply receive more excitatory drive compared to neighboring principal neurons, an idea consistent with the notion of STDP-dependent potentiation of synapses within the BLA during fear learning. However, this does not account for the lack of effect on EXT neurons, which receive enough excitatory input to express TetTag, and would presumably by innervated by similar CS-information-carrying inputs. An alternative explanation is that parameters not captured by static synaptic measurements, such as dynamic oscillatory phenomena, endow the PV-interneuron network with properties capable of modulating FC neurons and EXT neurons separately.

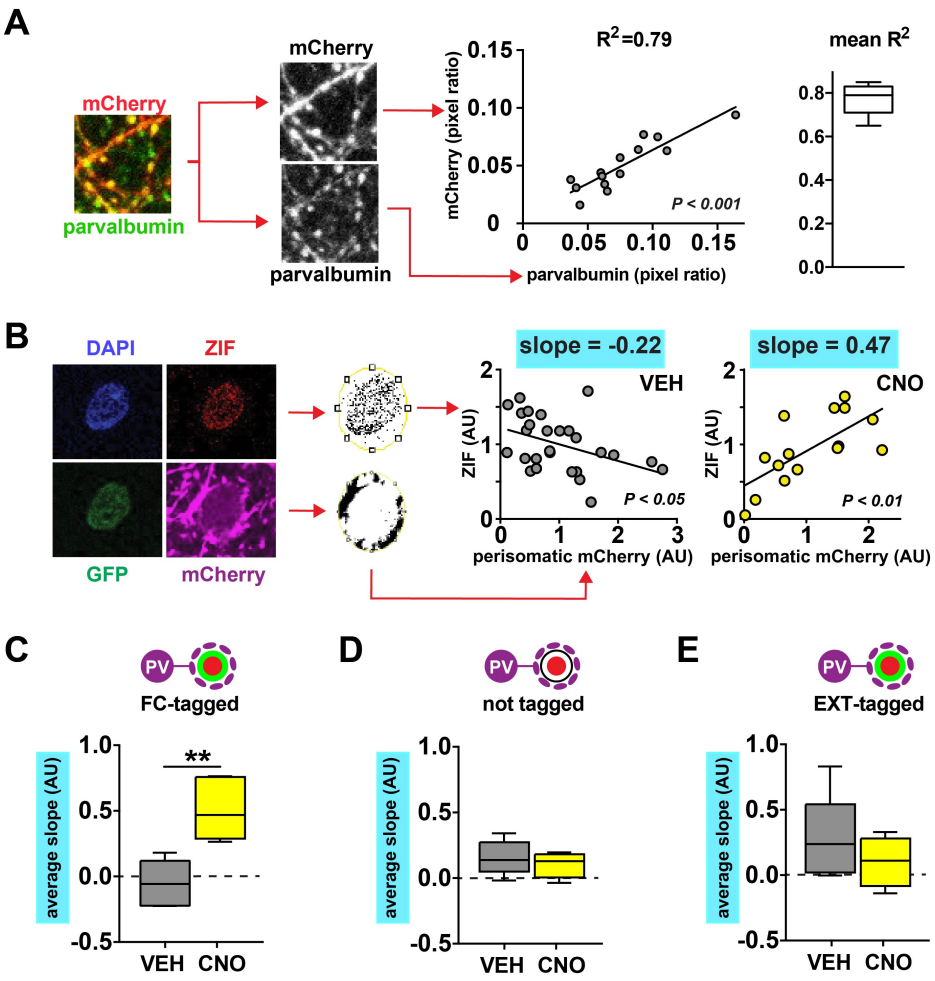

Figure 1.3. Structured perisomatic inhibition selectively silences BLA fear neurons during post-extinction retrieval.
A) Left: example of confocal image showing overlap of mCherry and parvalbumin immunofluorescence. Middle: example data from one mouse demonstrating high degree of correlation between the local presence of perisomatic mCherry and parvalbumin (linear regression: $F(1,12) = 46.16$, $P < 0.0001$, n = 14 neurons). Right: mean R^2 of 5 mice analyzed as in middle panel. **B)** Left: example GFP+ZIF+ neuron demonstrating detection of perisomatic mCherry signal. Middle: example data from a VEH injected mouse with a negative correlation between perisomatic mCherry and ZIF signal (linear regression: $F(1,25) = 5.106$, $P = 0.0328$, n = 27 neurons). Right: example data from a CNO injected mouse with a positive correlation between perisomatic mCherry and ZIF signal (linear regression: $F(1,13) = 11.83$, $P = 0.0044$, n = 15 neurons). **C)** CNO injection 30 min prior to post-extinction retrieval results in a positive relationship between perisomatic mCherry and ZIF expression in FC-tagged GFP+ neurons (unpaired t-test: $t(8) = 4.277$, $P = 0.0027$, VEH n = 5 mice, CNO n = 5 mice). **D-E)** CNO has no effect on the relationship between perisomatic mCherry and ZIF expression in GFP- neurons that were not tagged during fear conditioning (**D**, unpaired t-test: $t(9) = 0.798$, $P = 0.4454$, VEH n = 5 mice, CNO n = 6 mice), or in EXT-tagged GFP+ neurons (**E**, unpaired t-test: $t(8) = 1.06$, $P = 0.3199$, VEH n = 6 mice, CNO n = 4 mice).

To address this possibility, we recorded local field potentials (LFPs) from the BLA during our conditioning, extinction, and retrieval behavioral paradigm. This allowed us access to the emergent and dynamic properties of the BLA circuit, which we could follow across multiple learning trials as well as during PV-interneuron silencing (Figure 4A). We found, as previously reported, a prominent 3-6Hz oscillation in BLA following fear conditioning that was associated with freezing behavior(Figure 1.4 B) [74]. Consistent with this association, we found that extinction learning decreased the power of this oscillation. Interestingly, we found a separate oscillation in the 6-12Hz range that was more prominent in states of relative safety or exploration (Figure 1.4 B). We reasoned that, if these two bands represented distinct and antagonistic network states, the balance of them could determine the functional output of the BLA. In support of this, we found that silencing BLA PV-interneurons, thereby increasing fear behavior, increased the power of the 3-6Hz oscillation, decreased the power of the 6-12Hz oscillation, and increased the 3-6Hz/6-12Hz power ratio (Figure 1.4 H-K). In addition, the power ratio correlated with freezing levels across animals, again supporting the notion that the balance of these two opposing network states determines the functional output of BLA (Figure 4K). Interestingly, however, when analyzing a similar "high fear" state – post fear-conditioning but prior to extinction – we found that only the 3-6Hz band, not the 6-12Hz band, was affected by conditioning, and the 3-6Hz/6-12Hz ratio did not correlate as well with behavior across animals (Figure 1.4 D-G).

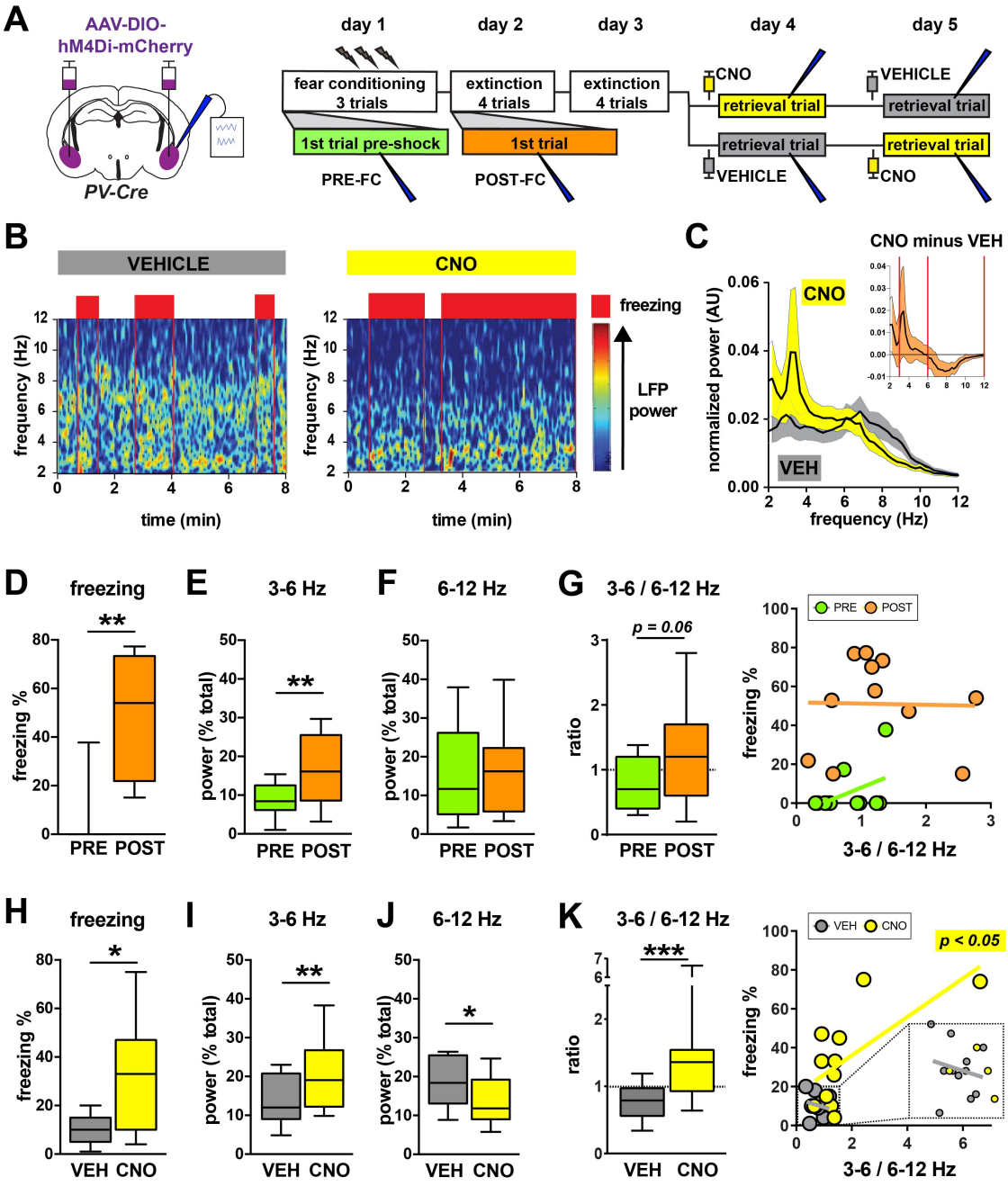
A
AAV-DIO-
hM4Di-mCherry
PV-Cre
day 1
day 2
day 3
day 4
day 5
fear conditioning
3 trials
extinction
4 trials
extinction
4 trials
1st trial pre-shock
1st trial
PRE-FC
POST-FC
CNO
retrieval trial
VEHICLE
retrieval trial
retrieval trial
VEHICLE
retrieval trial
CNO
B
VEHICLE
CNO
freezing
LFP
power
frequency (Hz)
time (min)
C
CNO minus VEH
CNO
VEH
normalized power (AU)
frequency (Hz)
D
freezing
freezing %
PRE POST
E
3-6 Hz
power (% total)
PRE POST
F
6-12 Hz
power (% total)
PRE POST
G
3-6 / 6-12 Hz
p = 0.06
ratio
PRE POST
PRE POST
freezing %
3-6 / 6-12 Hz
H
freezing
freezing %
VEH CNO
I
3-6 Hz
power (% total)
VEH CNO
J
6-12 Hz
power (% total)
VEH CNO
K
3-6 / 6-12 Hz
ratio
VEH CNO
VEH CNO
p < 0.05
freezing %
3-6 / 6-12 Hz

Figure 1.4. BLA PV-interneurons control the balance between two functionally opposed low frequency oscillations.
A) Experimental design: mice were infused bilaterally with AAV-Syn-DIO-hM4Di-mCherry and simultaneously implanted with recording electrodes in BLA. Mice were then subjected to contextual fear conditioning, extinction, and retrieval. **B)** Example of full trial spectrograms from a single animal demonstrating a 3-6 Hz oscillation during freezing (red boxes: periods of >50% freezing per bin), as well as a shift towards increased 3-6 Hz power compared to 6-12 Hz power caused by silencing BLA PV interneurons. **C)** Averaged normalized 2-12 Hz power spectra from CNO and VEH retrieval trials. INSET: Running difference between power spectra from CNO trials and VEH trials, calculated by subtracting averaged VEH trial spectra from CNO trial spectra (n = 11; red vertical lines mark 3-6 Hz and 6-12 Hz intervals used for quantification). **D-G)** Fear conditioning leads to increased freezing in the conditioned context (**D**, Wilcoxon matched-pairs: W = 64, P = 0.0020, n = 11 mice), increased 3-6 Hz power (**E**, paired t-test: t(10) = 3.481, P = 0.0059, n = 11 mice), no change in 6-12 Hz power (**F**, paired t-test, t(10) = 0.3747, P = 0.7157, n = 11 mice), and caused a trend for an increased 3-6 / 6-12 Hz power ratio (**G left**, paired t-test: t(10) = 2.074, P = 0.0648). The 3-6 / 6-12 Hz power ratio does not correlate with freezing before (**G right PRE**, linear regression: F(1,9) = 2.299, P = 0.1638, n = 11 mice), or after fear conditioning (**G right POST**, linear regression: F(1,9) = 0.003989, P = 0.9510, n = 11 mice). **H-K)** Silencing PV-interneurons during post-extinction retrieval leads to increased freezing in the conditioned context (**H**, paired t-test: t(10) = 2.867, P = 0.0167, n = 11 mice), increased 3-6 Hz power (**I**, Wilcoxon matched-pairs: W = 56, P = 0.0098, n = 11 mice), decreased 6-12 Hz power (**J**, paired t-test: t(10) = 2.427, P = 0.0356, n = 11 mice), and an increased 3-6 / 6-12 Hz power ratio (**K left**, Wilcoxon matched-pairs: W = 66, P = 0.0010, n = 11 mice). The 3-6 / 6-12 Hz power ratio correlates with freezing in CNO injected mice (**K right CNO**, linear regression: F(1,9) = 7.423, P = 0.0234, n = 11 mice), but not in VEH injected mice (**K right VEH**, linear regression: F(1,9) = 0.3594, P = 0.5636, n = 11 mice).

Given our previous data that A) extinction learning induces structural plasticity in the BLA PV-network and B) silencing PV-interneurons after, *but not before*, extinction learning, selectively impaired the ability to suppress conditioned fear, we reasoned that extinction learning itself reorients the PV network such that the two network states are anti-correlated, thereby allowing the 6-12Hz band to act as a "safety" signal capable of suppressing the 3-6Hz signal. To investigate this further, we calculated and cross-correlated instantaneous amplitude of the two frequency bands to test for any relationship between them (Figure 1.5 A-B). We found, interestingly, that the signals exhibited a negative correlation only after extinction learning, consistent with the notion that extinction learning alters the relationship between the two bands (Figure 1.5 C-D).

Strikingly, this relationship switches to a positive correlation after silencing of PV-interneurons, indicating that the BLA PV-network is critical for mediating the competitive interaction between the 3-6Hz and 6-12Hz bands following extinction learning (Figure 1.5 C-D). Interestingly, we also found that the competition between the two bands is dynamic, as indicated by its rapid appearance around the onset of freezing behavior (data not shown). Taken together, these data are consistent with the intriguing notion that extinction learning modifies the BLA PV-network to allow competition between two distinct oscillations, resulting in the suppression of the 3-6Hz "pro-fear" band and an overall reduction in associated fear behavior.

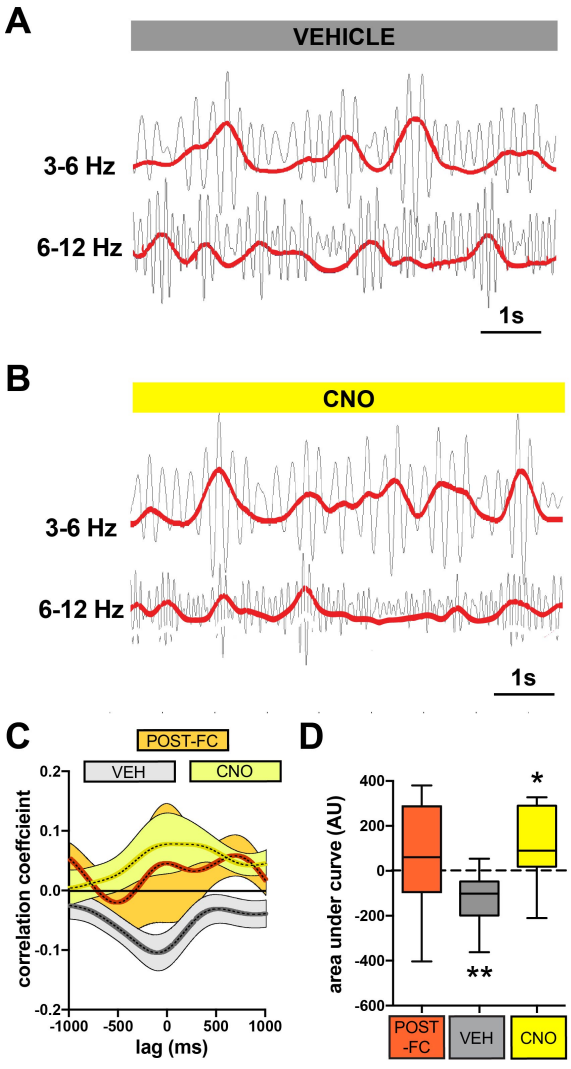

Figure 1.5. BLA PV-interneurons enable competition between 3-6 Hz and 6-12 Hz oscillations.
A-B) Examples of 3-6 Hz and 6-12 Hz filtered traces indicate a negative correlation between the instantaneous powers of these two bands in a vehicle-injected mouse, and a positive correlation in a CNO-injected mouse. **C-D)** Averaged cross-correlograms and quantified area under curve (AUC) were calculated for POST-FC, VEH, and CNO conditions (see experimental design in Fig. 3A), and the presence of a significant positive or negative correlation was tested by comparing the AUC values with a theoretical median of zero. This revealed a significant negative correlation between the 3-6 Hz and 6-12 Hz bands after extinction in VEH injected mice (**D VEH**, Wilcoxon Signed Rank Test with theoretical median = 0: W = -51, P = 0.0059, n = 11 mice), which is shifted to a significant positive correlation by CNO injection (**D CNO**, Wilcoxon Signed Rank Test with theoretical median = 0: W = 46, P = 0.0420, n = 11 mice). There was no significant correlation before extinction (**D POST-FC**, Wilcoxon Signed Rank Test with theoretical median = 0: W = 17, P = 0.3594, n = 9 mice).

One prediction of the above model would be that exogenous induction of 8Hz activity in the BLA would be capable of suppressing both the 4Hz oscillatory activity, as well as reducing pro-fear BLA output and thus reducing overall freezing levels. Critically, one would predict this effect would only be seen *after* extinction learning, representing the new network structure induced by the learning experience. In order to test this hypothesis, as well as well as to directly test the causal role of BLA oscillations in the production of fear behavior, we employed an optogenetic approach capable of driving oscillatory activity in BLA via the manipulation of PV-interneurons. Previous work from the Herry lab has demonstrated that application of 4Hz sinusoidal optical input onto PV-interneurons was capable of generating 4Hz oscillatory activity in dorsal mPFC and producing fear behavior[58]. To employ a similar "gain-of-function" approach, we injected AAV-Syn-DIO-Chr2-mCherry virus unilaterally into the BLA of PV-Cre mice and implanted optical cannulas in the BLA, thus exerting fine spatiotemporal control over BLA PV-interneuron activity (Figure 1.6 A-B). Mice were subjected to our conditioning and extinction paradigm as above, and optical stimulation at 4Hz or 8Hz was performed at four separate points: during the habituation phase of the first conditioning trial, during retrieval post-conditioning, during retrieval post-extinction learning, and during exposure to a neutral context (Figure 1.6 C). Thus, we were not only able to address the causal role of BLA PV-interneurons in producing functional oscillations and the role of these oscillations in controlling BLA output, but we were also able to test the necessity of learning for these effects, thereby probing how fundamental BLA network properties are modified by experience.

We found that rhythmic stimulation of the BLA PV-network at 4Hz or 8Hz in a novel context before fear conditioning had no effect on freezing levels (Figure 6D). Similarly, stimulation after fear conditioning, but before extinction learning, produced no observable behavioral effect (Figure 1.6 E). Strikingly, however, stimulation at 4Hz and 8Hz following extinction learning produced consistent augmentation, and suppression, respectively, of fear behavior. In other words, only after extinction learning does rhythmic driving of BLA activity exert bidirectional control over conditioned fear behavior (Figure 1.6 F). To test whether this was due to retrieval of a specific conditioned response, or a general non-specific effect of stimulation, we tested the same animals in an unconditioned, neutral context (context B). Stimulation of neither 4Hz nor 8Hz produced a behavior effect in these animals, indicating that the observed effect is specific to the conditioned context (Figure 1.6 G). Given this specificity, and the dependence on previous learning, we can conclude that 4Hz and 8Hz stimulation bidirectionally affect the balance of the retrieval of two opposing memory traces; a conditioned fear memory and the subsequent extinction memory. These experiments also demonstrate conclusively a causal relationship between the two oscillations and behavioral output. However, because BLA does not directly generate behavior, other circuits must be involved and therefore must participate in and/or be receptive to oscillatory BLA network activity. Additionally, the mechanisms underlying the mysterious finding that the behavioral effect of 4 or 8Hz rhythmic induction depends on previous learning remain unclear. I will address both of these issues in Section II.

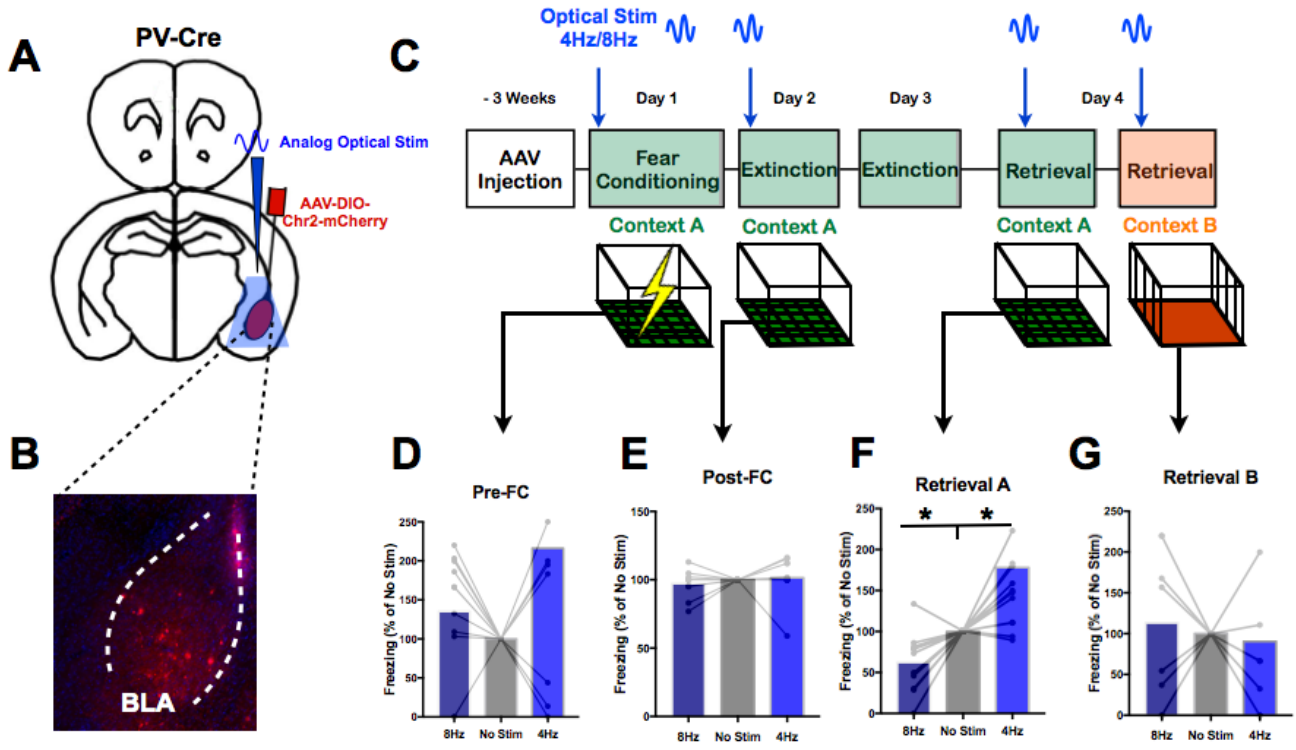

Figure 1.6: Analog stimulation of BLA PV-interneurons modulates fear memory retrieval in a bidirectional, learning-dependent manner
A) Injection and recording strategy. AAV-DIO-Chr2-mCherry is injected unilaterally into the BLA of PV-IRES-Cre mice. Optical cannulas are placed into the BLA of injected animals, allowing frequency-specific analog stimulation of BLA PV-interneurons. **B)** Example image of Chr2-mCherry expression in BLA PV-interneurons. **C)** Experimental design. After virus injection and electrode/cannula implantation, mice were subjected to a modified version of our standard behavioral design, including fear conditioning, extinction, and retrieval trials. Optical stimulation occurred at the onset of the first FC trial (Pre-FC), during exposure to the fear conditioned context, prior to extinction learning (Post-FC), during exposure to that same context after extinction learning (Retrieval A), and exposure to a novel, neutral context (Retrieval B). **D-G)** 4Hz or 8Hz analog stimulation results in increased or decreased freezing, respectively, only after extinction learning and specifically in the conditioned context. Pre-FC P = 0.71, 0.22; Post-FC P = 0.99, 0.98; Retrieval A P = 0.01, 0.02; Retrieval B P = 0.99, 0.99; n = 11, Paired-tests, Dunnett's multiple comparison's correction.

An additional remaining question from the above studies is how the role of PV-interneurons in mediating selective FC neuron ensemble suppression and their role in facilitating interactions between higher-level functional network states are related to one another. In other words, what connects the two ensembles (FC and EXT neurons) and the two network states (3-6Hz and 6-12Hz), and what explains the critical role of PV-interneurons in modulating both following extinction learning? An intriguing explanation lies in a phenomenon common to many physical systems known as resonance. Simply put, resonance is a property of preferential frequency response. Anyone who has ever sung in the shower knows that certain notes sound louder than others, even given constant vocal amplitude. This is due to the fact that the spatial constraints of the room produce an acoustic profile such that certain frequencies (tones) resonate, resulting in preferential amplification or dampening. Neurons are also capable of exhibiting resonance, which is easily measured by applying a sinusoidal current of increasing frequency and constant amplitude into the cell, and measuring its voltage response. If a neuron has a preferred (resonant) frequency, it will have a higher amplitude voltage response at that frequency compared to others [83]. With this notion in mind, we decided to test whether discrete functional ensembles in the BLA exhibit resonance, and further whether their resonance profiles can partially explain the parallel lines of evidence connecting our ensemble and oscillatory data.

To test whether functional BLA ensembles exhibit differential resonance, we subjected TetTag animals to our fear conditioning and extinction paradigm as previously in order to produce "FC-Tagged" and "EXT-Tagged" neurons (See Methods). We then analyzed the intrinsic resonance properties of these neurons in the general manner described above (Figure 1.7 A-B, see Methods for more detail) [83]. We hypothesized that neurons active during fear learning would be tuned preferentially to 4Hz frequency stimulation, whereas neurons active during extinction learning would be tuned to higher frequencies. We found, as expected, that "FC-Tagged" neurons preferentially responded to sinusoidal current injection at 4Hz compared to 8Hz, consistent with previous reports of BLA tuning. Interestingly, GFP- "non-ensemble" neurons which were not active during fear learning, had a less pronounced tuning curve, as did "EXT-Tagged" neurons, suggesting they are capable of responding equally well to higher frequency oscillatory input (Figure 1.7 C). This raises the intriguing possibility that a potential effect of BLA oscillations at 4Hz or 8Hz would be the preferential recruitment of functionally distinct or even opposing pyramidal neuron ensembles based on the resonant properties of the participant neurons. In order to directly test this hypothesis, we measured both intrinsic resonance and firing rate of BLA pyramidal neurons during sinusoidal optical stimulation of BLA PV-interneurons at 4 or 8Hz (Figure 1.7 A,D, see methods). This way we could partially mimic the microcuitry effect of an oscillation *ex vivo.* Though our data are thus far preliminary, they do suggest that, interestingly, neurons that are tuned preferentially to 4Hz input will preferentially during an ongoing 4Hz oscillation (generated by 4Hz PV stimulation) compared to during an 8Hz oscillation (Figure 1.7 D-E). Taken together, these data suggest that PV-interneurons are a critical hub in controlling the output of

distinct functional pyramidal neuron ensembles through a combination of oscillatory activity and matched resonance of the target ensembles.

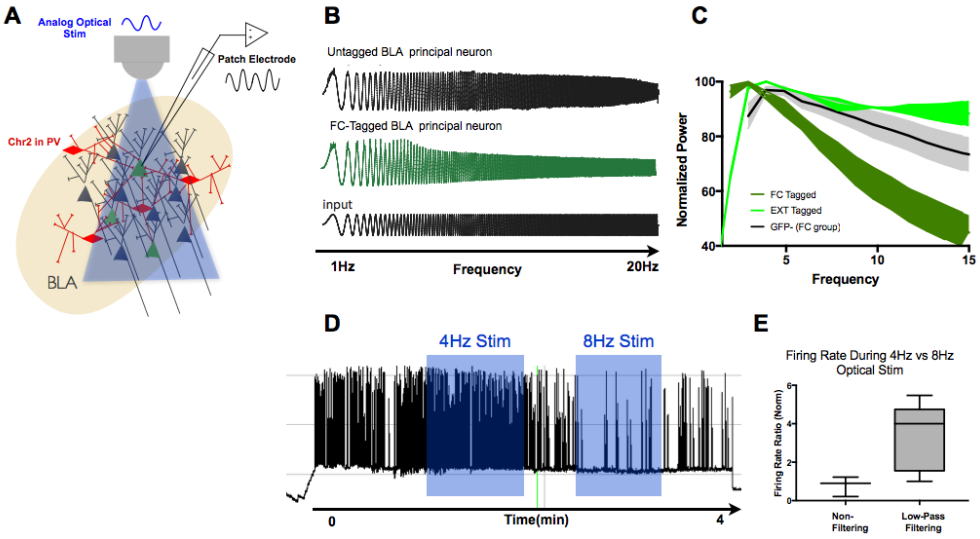

Figure 1.7: Divergent intrinsic resonance properties of BLA principal neurons may define functional specificity by affecting participation in opposing rhythms
A) Experimental design. TetTagged and untagged BLA principal neurons are recorded from using patch electrophysiology in acute BLA slices. Instrinsic resonance properties of neurons are determined by applying sinusoidal subthreshold current to neurons and measured voltage response **(**see example in panel B). Participation in rhymic output is inferred from firing rate of BLA neurons during analog optical stimulation of local PV-interneurons at 4 or 8Hz. **B)** Example principal neurons demonstrating a non-resonant, untagged BLA principal neuron (top trace), and a resonant, low-pass filtering "FC-Tagged" BLA principal neuron (middle trace)**.** **C)** Averaged FFT from instrinsic resonance measurements across FC-tagged, EXT-Tagged, and untagged BLA principal neurons. FC-tagged neurons may have more pronounced low-pass filtering properties than other functional subsets within the BLA. **D)** Example data from a BLA principal neuron demonstrating differential firing rate during 4 or 8Hz analog optical stimulation of PV-interneurons. **E)** Summary data after classifying neurons as either "non-resonant" or "low-pass filtering" suggesting that intrinsic resonance properties may determine firing rate of BLA principal neurons during 4 or 8Hz stimulation of PV-interneurons. ALL DATA PRELIMARY, no statistical analysis performed. Data generated in collaboration with Jamie Maguire.

In summary, our findings to this point have demonstrated a critical role for PV-interneurons in the suppression of conditioned fear behavior following extinction learning. They achieve this suppression by mediating the selective suppression of fear-encoding neurons, and not other pyramidal neuron ensembles, within the BLA. The circuit properties responsible for this learning-dependent role are not fully explained by the amount of perisomatic inhibition received by pyramidal neurons, and require an appreciation of the dynamics of circuit function. Towards this end, we have identified two distinct oscillatory states of BLA, one defined by prominent 4Hz oscillations and freezing behavior, and the other by 8Hz oscillations and safety behavior. PV-interneurons appear to control the interaction of these two network states (slope of energy diagram), as loss of PV-interneuron activity by chemogenetic silencing altered the sign of the relationship between them. We have provided some evidence that the relationship between the network state and ensemble activation may rely upon differences in the resonance properties among distinct BLA ensembles, such that they bias inclusion into an ongoing oscillatory state and therefore would critically modulate BLA output. Finally, we have established that optogenetic activation of PV-interneurons at 4 or 8Hz resulted in learning-dependent, bidirectional changes to fear behavior following extinction learning. There can be no doubt that extinction learning confers the BLA PV-network with a fundamental role in suppressing fear behavior, and indeed in toggling between fear and safety states, and our identification of both cellular and oscillatory correlates of this feature will be informative for future studies concerning extinction learning and amygdala function. Still, though, some questions remain: 1) What effect do BLA PV-interneurons have on circuit *output* and thus how do they exert influence over behavior?

2) What accounts for the learning-dependency of PV-interneuron optogenetic manipulation? In order to answer these and other questions, we must include other aspects of the circuit outside of the BLA.

Section 2: BLA PV-interneurons route information transfer across the mPFC-BLA circuit to controls fear behavior in a learning-dependent manner

Our findings to this point identify PV-interneurons within the BLA as a critical regulator of both functional neuronal ensembles and oscillatory network states, as well as substrates of extinction learning with BLA that critically depend on PV-interneurons. As mentioned in the introductory chapter, however, BLA PV-interneurons are known to receive long-range input from several brain regions, consistent with a putative role in effecting feedforward inhibition [61]. Indeed, previous studies have demonstrated nicely that PV-interneurons in cortex can mediate long-range feedforward inhibition onto principal cells [58,84]. Theoretically, the combination of feedforward inhibition and the feedback/lateral inhibitory motifs discussed in the previous section would have the capacity to dynamically regulate interregional communication. As previous studies have found modulation in mPFC-BLA communication to be essential for fear and extinction behavior (cite), we decided to investigate the possibility of feedforward inhibitory modulation of BLA circuitry through direct mPFC→BLA(PV) input. To do so, we employed a Cre-dependent trans-synaptic labeling strategy, which allowed us to specifically label presynaptic inputs onto genetically defined subsets of neurons [85]. We found input onto BLA PV-interneurons from several regions, including mPFC. Interestingly, we found more input from infralimbic cortex (ILC), a subdivision of mPFC

known to be essential for extinction learning compared to prelimbic cortex, known to be essential for fear learning and freezing behavior(Figure 2.1 A) [86]. To test whether this was simply a result of denser input from ILC→BLA generally, or whether it was a specific feature of ILC→BLA (PV) connection, we injected mice with cholera-toxin subunit B, which nonselectively labels somata with projections to the target area. Interestingly, we found, in contrast to our mPFC→BLA(PV) data, more labeled somata in PLC compared to ILC, indicating preferential engagement of PV-interneurons by ILC projection neurons (Figure 2.1 B). This finding is generally consistent with a previous report demonstrating denser projections from PLC to BLA compared to ILC [87]. Our findings, however, raise the intriguing possibility that while ILC→BLA projections are sparser in general, their preferential engagement of the BLA PV-network may be essential for extinction learning and subsequent fear suppression. This would be consistent with both our work demonstrating the necessity of the BLA PV-network following extinction learning and the known role of ILC in extinction [88–91].

One of the major targets of BLA projection neurons is mPFC [31,35,84]. Indeed, previous work has implicated preferential engagement of PLC or ILC in functional delineations of BLA ensembles [36]. In addition, labeling of BLA projection neurons revealed dense, layer-specific innervation of mPFC from BLA (Figure 2.1 C). This fact, together with our work, raises the possibility of a functional reciprocal circuit between mPFC and BLA. To directly test for this, employed a similar Cre-dependent rabies strategy as above, but rather than using genetically defined Cre mouse line, we conferred specificity of labeling by injecting a retrogradely infecting Cre-expressing virus into mPFC. Thus, by using the Cre-dependent transsynaptic strategy in BLA, we could

exclusively label inputs onto BLA projections that project to mPFC. We found dense input from multiple regions, including ventral hippocampus (data not shown). We also found, as predicted, labeled somata in mPFC, indicating the presence of a mPFC→BLA→mPFC disynaptic loop (Figure 2.1 B). Taken together, these data indicate that BLA PV-interneurons participate in a reciprocal mPFC-BLA circuit, and implicate the BLA PV-network in modulating activity not only within BLA, but also across the broader circuit.

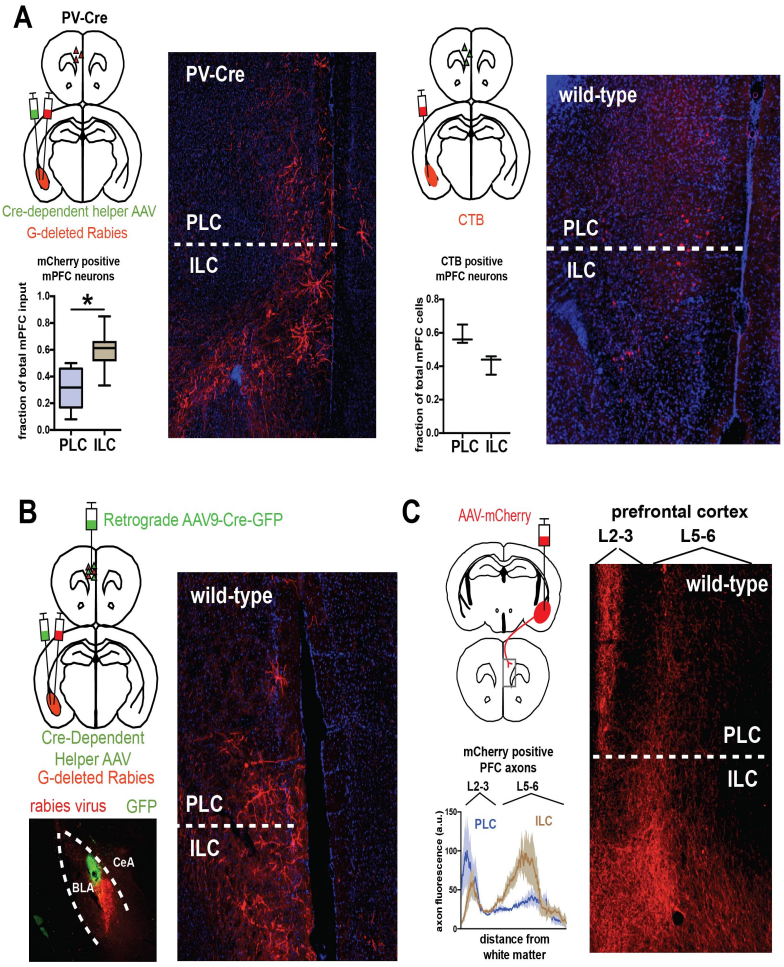

Figure 2.1. BLA PV-interneurons participate in a reciprocal BLA-mPFC circuit.
A) Left: to trace monosynaptic inputs on BLA PV-interneurons, helper virus (AAV9-Ef1α-FLEX-GTB) and mCherry-expressing rabies virus was injected into BLA of PV-Cre mice. The example image shows mCherry-positive neurons in mPFC that synapse onto BLA PV-interneurons, with the graph showing that a larger fraction of mCherry-positive neurons was located in the ILC versus PLC (**A left**, paired t-test: $t(7) = 2.8$, $P = 0.0265$, $n = 8$ mice). Right: to trace all inputs to the BLA, CTB was injected into the BLA of wildtype mice. The example image shows CTB-positive neurons in the mPFC that project to the BLA, with the graph showing similar fractions of CTB-positive neurons located in the PLC and ILC (**A right**, paired t-test: $t(2) = 2.463$, $P = 0.1328$, $n = 3$ mice). **B)** To selectively label monosynaptic inputs onto BLA→mPFC projecting neurons, retrograde AAV9-Cre-GFP was injected into mPFC, and helper virus and mCherry-expressing rabies virus injected into BLA three weeks later. Bottom left: example image of injection site of mCherry-expressing rabies virus (red) in the BLA and green nuclei from retrograde AAV9-Cre-GFP virus injected into mPFC. Right: example image demonstrating mCherry-positive neurons in mPFC that synapse onto BLA→mPFC projecting neurons. **C)** To label BLA→mPFC projections, AAV-Cre-GFP was injected into mPFC, and AAV-Syn-DIO-hM4Di-mCherry was injected into BLA. Right and bottom left: representative image and quantification showing dense innervation of superficial PL layers (L2-3) and deep IL layers (L5-6) by BLA projection neurons ($n = 4$).

To investigate the above possibility directly, we compared the LFP activity in mPFC following BLA PV-interneuron silencing. We found, as in BLA, an increase in 3-6Hz:6-12Hz power ratio, indicating that the changes in BLA activity caused by silencing PV-interneurons were mirrored in mPFC (Figure 2.2 B-D). To investigate whether this had any effect on mPFC ensemble activity, we again used TetTag to label populations of neurons active during learning, and quantified the effect BLA PV-interneuron silencing on their reactivation state. We found that silencing BLA PV-interneurons increased the reactivation of TetTagged neurons in superficial PLC, and that the percentage of reactivation correlated with freezing behavior during retrieval, consistent with the proposed role of BLA→PLC projections in fear expression (Figure 2.2 F-G) [35,36]. In contrast, we observed a decrease in the reactivation of deep-layer ILC TetTagged neurons, consistent with their proposed role in mediating fear suppression and/or extinction learning (Figure 2.2 I) [36]. To investigate the functional role of these mPFC ensembles, something which has not yet been done using TetTag, we performed a separate experiment comparing the labeling of ensembles within mPFC by TetTag during fear-conditioning (FC group), exposure to a novel context without footshock (NS group), or while the animals were left in their homecage (HC group). We also compared the reactivation of these neurons when put back in the conditioning/novel context (with the exception of the HC group). We found that, consistent with their proposed role in fear behavior, neurons in PLC were preferentially TetTagged in the FC group (Figure 2.2 H). In contrast, we found that neurons in ILC were preferentially tagged in the NS group (Figure 2.2 J). It is possible that these neurons are therefore more active when an animal

is actively learning that a context is safe, such as during exposure to a novel context, or to a previously conditioned context that is no longer paired with an aversive stimulus (footshock). This would be consistent with a role in extinction learning, during which an animal has to learn that a previously conditioned context is now safe. In support of this, we found that neurons in ILC were preferentially reactivated following retrieval in the FC group (data not shown). Thus, their activity in these two states (initial exposure to a novel context and a neutral experience in a previously conditioned context) may be characterized by a modified version of the phenomenon known as prediction error (PE), which I will discuss further in the discussion. For now, suffice it to say that the ensembles in PFC and ILC likely perform broadly antagonistic functions within the brain with regard to fear behavior, and our data indicate that they are regulated by BLA output (which is, in turn, regulated by BLA PV interneurons).

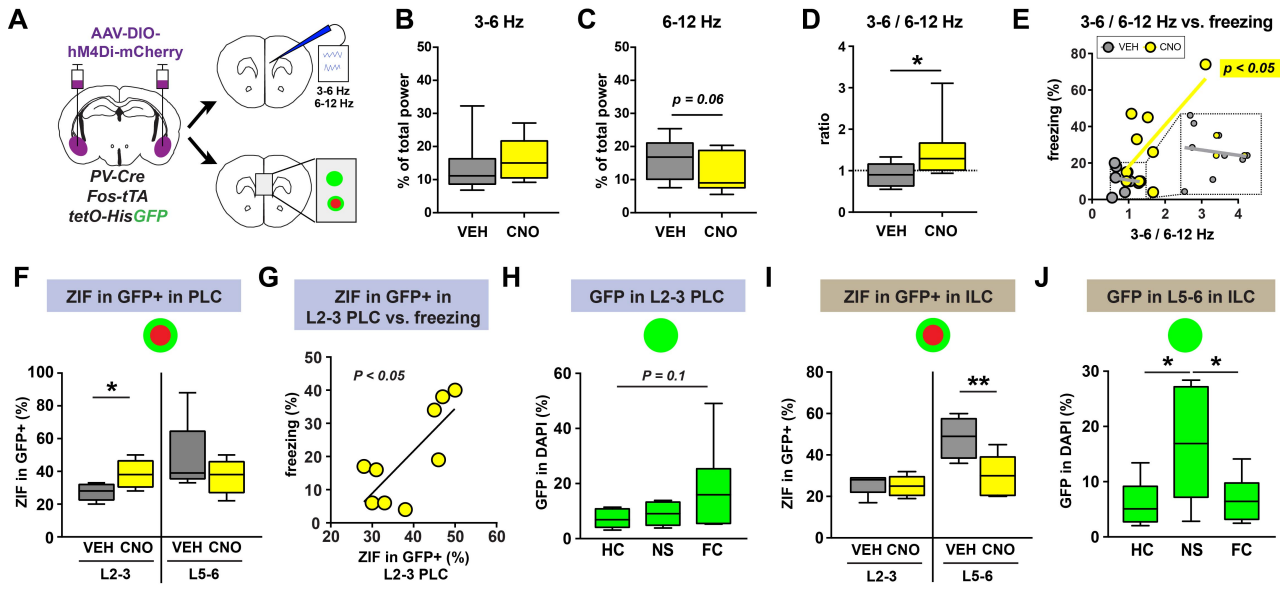

Figure 2.2. Silencing BLA PV-interneurons shifts mPFC ensemble dynamics towards a pro-fear state.

A) Chemogenetic silencing of BLA PV-interneurons was combined with either *in vivo* LFP recording in the mPFC using the same group of mice as in figure 3, or with TetTag analysis of the mPFC using the same group of mice as in figure 1. **B-D)** Silencing PV-interneurons during post-extinction retrieval increased the 3-6 / 6-12 Hz power ratio in the mPFC (**D**, Wilcoxon matched-pairs: W = 39, P = 0.0195, n = 9 mice), without significantly changing the 3-6 Hz power (**B**, Wilcoxon matched-pairs: W = 23, P = 0.2031, n = 9 mice), or the 6-12 Hz power (**C**, paired t-test, t(8) = 2.201, P = 0.0589, n = 9 mice). **E)** The 3-6 / 6-12 Hz power ratio correlated with freezing in CNO injected mice (linear regression: F(1,7) = 5.673, P = 0.0488, n = 9 mice), but not in VEH injected mice (linear regression: F(1,7) = 0.1333, P = 0.7258, n = 9 mice). **F)** Silencing BLA PV interneurons during post-extinction retrieval increased the reactivation of superficial PLC GFP+ neurons (L2-3, unpaired t-test, t(12) = 2.689, P = 0.0197, VEH n = 5 mice, CNO n = 9 mice), but had no effect on deep PLC GFP+ neurons (L5-6, Mann-Whitney test: U = 17, P = 0.4935, VEH n = 5 mice, CNO n = 9 mice). **G)** The percentage of reactivated GFP+ neurons in the superficial PLC correlated with freezing behavior (linear regression: F(1,7) = 9.977, P = 0.0160, n = 9 mice). **H)** Fear conditioning caused a trend for increased GFP+ tagged neurons in the superficial PLC as compared to mice left in the homecage (one-way ANOVA: F(2,17)=2.684, P = 0.0970, homecage/HC n = 8 mice, no-shock/NS n = 5 mice, fear conditioned/FC n = 7 mice; Tukey's multiple comparisons test HC versus FC: q(17) = 3.139 , P = 0.0964). **I)** Silencing BLA PV interneurons during post-extinction retrieval decreased the reactivation of deep ILC GFP+ neurons (L5-6, unpaired t-test: t(12) = 3.336, P = 0.0059, VEH n = 5 mice, CNO n = 9 mice), but had no effect on superficial GFP+ neurons (L2-3, Mann-Whitney test: U = 20, P = 0.7962, VEH n = 5 mice, CNO n = 9 mice). **J)** Deep layer ILC neurons are preferentially tagged in animals exposed to a novel context without shock, as compared to homecage and fear conditioned animals (one-way ANOVA: F(2,19) = 5.282, P = 0.0150, homecage/HC n = 8 mice, no-shock/NS n = 6 mice, fear conditioned/FC n = 8 mice; Tukey's multiple comparisons test HC versus NS: q(19) = 4.209, P = 0.0202; Tukey's multiple comparisons test NS versus FC: q(19) = 3.921, P = 0.0311).

Our findings demonstrate that BLA PV-interneurons are critical for the regulation of both mPFC oscillatory and functional ensemble activity. Silencing BLA PV-interneurons shifts mPFC activity towards a more pro-fear state, as characterized by an increased 3-6Hz:6-12Hz power ratio. In addition, silencing PV-interneurons increases the reactivation of fear-promoting PLC ensembles, while reducing the activity of learned safety or extinction-encoding ILC ensembles. Taken together our tracing data, the above findings implicate BLA PV-interneurons in both mPFC→BLA interaction, and in mediating the flow of information from BLA to mPFC. Due to this reciprocal nature of the mPFC-BLA circuit, as well as the above phenomena, we hypothesized that PV-interneurons may act as a router for functional communication between mPFC and BLA, specifically within the carrier frequency bands delineated in the previous section. To test this directly, we analyzed animals from which we had recorded LFPs simultaneously from both mPFC and BLA and calculated coherence and directionality of mPFC-BLA oscillations in the 3-6Hz and 6-12Hz bands. We found that silencing BLA PV-interneurons shifted peak coherence toward the 3-6Hz range, indicating that functional communication between the two regions is regulated by PV-interneurons (Figure 2.3 A). By calculating the directional (or lead-lag) relationship, one can begin to infer causal relationships between activity in the two regions, although several caveats apply (Figure 2.3 B) [92]. We found, as previously reported, that freezing following fear conditioning was associated with increased mPFC→BLA in the 3-6Hz range, suggesting that mPFC drives BLA activity in this "pro-fear" band following conditioning (data not shown) [74].

Interesting, following extinction learning, we found that the onset of freezing was associated with increased BLA→mPFC directionality in this band, indicating that the return of fear behavior following extinction learning is marked by a signal from BLA to mPFC in the 3-6Hz range (Figure 2.3 C-D). This was not true of the 6-12Hz band, which exhibited the opposite change at the onset of freezing, indicating that these two bands are regulated by distinct mechanisms and again indicating that they have opposite effects on behavioral output. Finally, we found that silencing PV-interneurons during retrieval increased the likelihood of BLA→mPFC directionality in the 3-6Hz band, and decreased BLA→mPFC directionality in the 6-12Hz band, demonstrating that PV-interneurons are a critical hub for regulating the directional communication of functional information across the mPFC-BLA circuit (Figure 2.3 E).

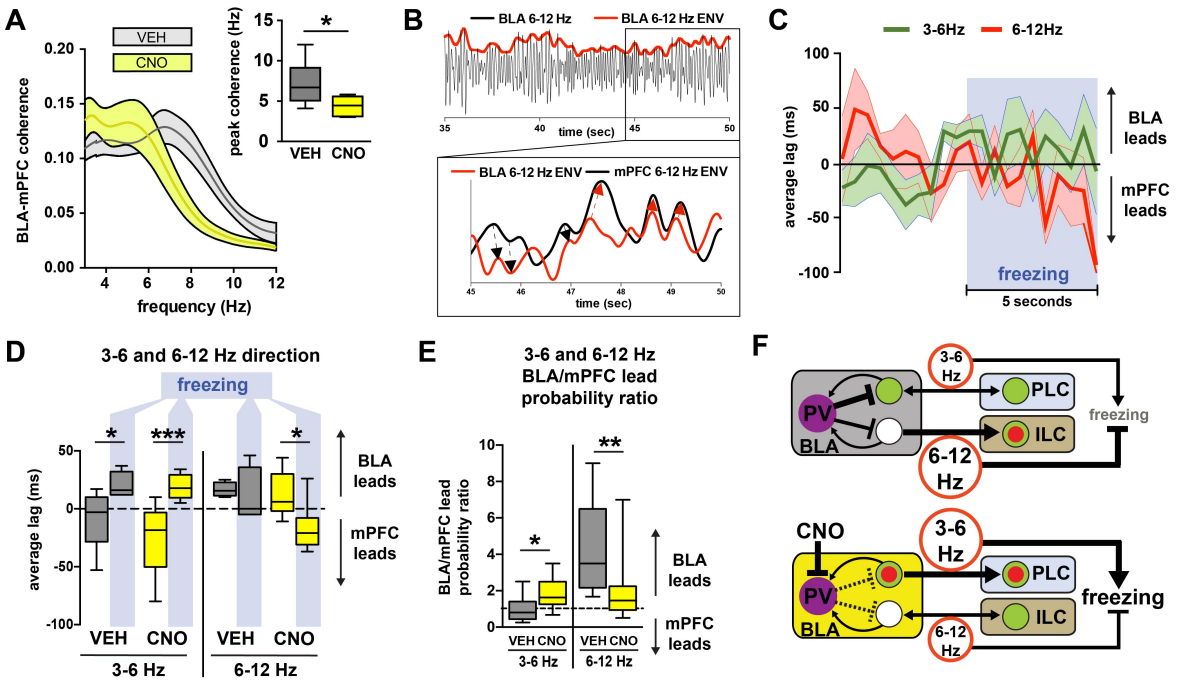

Figure 2.3. BLA PV-interneurons control directionality of 3-6 and 6-12 Hz oscillations following extinction.
A) Silencing BLA PV-interneurons during post-extinction retrieval shifts BLA-mPFC peak coherence towards the 3-6 Hz range (paired t-test: t(8) = 2.486, P = 0.0378, n = 9 mice). **B)** Example of a 6-12 Hz filtered trace demonstrating the extraction of instantaneous power envelopes (top) which can be cross-correlated (bottom) to determine leading (black arrows) or lagging (red arrows) relationships between two signals. **C)** Averaged lags from instantaneous power cross-correlations for both 3-6 Hz and 6-12 Hz bands during the 5 seconds preceding and the 5 seconds following the onset of freezing during retrieval trials (VEH n= 6 mice, CNO n = 8 mice, 1 freezing event per mouse per trial). **D)** The first 5 seconds following the onset of freezing is characterized by a shift towards a BLA lead for 3-6 Hz in both VEH and CNO groups (repeated measures two-way ANOVA: freezing F(1,11) = 29.12, P = 0.0002, CNO F(1,11) = 1.013, P = 0.3357, freezing x CNO F(1,11) = 1.411, P = 0.2598, VEH n = 5 mice, CNO n = 8 mice; Sidak's multiple comparisons test: pre-freeze vs freeze for VEH t(11) = 2.682, P = 0.0422; Sidak's multiple comparisons test: pre-freeze vs freeze for CNO t(11) = 5.308, P = 0.0005), and a shift towards a mPFC lead for 6-12 Hz in the CNO group (repeated measures two-way ANOVA: freezing F(1,9) = 5.487, P = 0.0439, CNO F(1,9) = 2.263, P = 0.1668, freezing x CNO F(1,9) = 2.355, P = 0.1592, VEH n = 4 mice, CNO n = 7 mice; Sidak's multiple comparisons test: pre-freeze vs freeze for VEH t(9) = 0.5063, P = 0.8593; Sidak's multiple comparisons test: pre-freeze vs freeze for CNO t(9) = 3.215, P = 0.0210). **E)** Silencing PV interneurons during post-extinction retrieval increased the probability of BLA leading mPFC in the 3-6 Hz band (paired t-test, t(9) = 2.303, P = 0.0468, n = 10 mice), while decreasing the probability that it will lead in the 6-12 Hz band (Wilcoxon matched-pairs: W = -45, P = 0.0039, n = 9 mice). **F)** Model based on our data and previous studies. Top: after extinction learning, BLA PV-interneurons suppress conditioned freezing and fear ensemble activation by enabling a 6-12 Hz oscillation to out-compete a 3-6 Hz oscillation throughout the BLA-mPFC circuit. Bottom: chemogenetic silencing of BLA PV-interneurons causes dysfunctional competition between the 6-12 and 3-6 Hz oscillations, which leads to an increase in 3-6 Hz power and BLA→mPFC directionality, an increase in the activation of fear ensembles in BLA and PLC, and an increase in conditioned freezing.

The above findings indicate that the return of fear following extinction learning is marked by a distinct 3-6Hz oscillatory signal from BLA to mPFC, the downstream interpretation and effect of which is itself regulated by learning. To directly test this model, we analyzed LFPs recorded simultaneously from mPFC and BLA during optogenetic manipulation of BLA PV-interneurons. We stimulated with a sine-wave function at 4Hz or 8Hz prior to conditioning (Pre-FC), during retrieval of conditioning (Post-FC), during retrieval following extinction learning (Post-FC-EXT), and during retrieval in a neural, unconditioned context (Ctx B). To reiterate our previous findings, stimulation of BLA PV-interneurons at 4Hz increased freezing levels only after fear conditioning and extinction learning, and exclusively in the conditioned context (Figure 6). Similarly, stimulation of BLA PV-interneurons at 8Hz decreased freezing levels only after extinction learning, and, once again, exclusively in the conditioned context. To account for this remarkable memory-specific effect, we hypothesized that the ability of the BLA to entrain mPFC activity is modified by learning, such that manipulation of BLA activity would only be capable of inducing synchronized activity in mPFC following learning. Specifically, we hypothesized that the manipulation of BLA PV-interneurons would be capable of inducing 4Hz synchronization following, but not before, fear conditioning, and similarly, would be able to induce 8Hz synchronization following, but not before, extinction learning. To test this, we measured phase coherence across the mPFC-BLA circuit during 4Hz stimulation, 8Hz stimulation, and no stimulation conditions. Phase coherence (referred to hereafter as simply coherence) is a measure of the degree of phase coupling between oscillations in two different regions.

High coherence indicates that oscillations in the two brain regions are phase coupled to one another and are thus more likely to a) arise from common mechanisms and b) result in functional output modulation [72]. This is due to the fact that oscillations may synchronize the output of one area (at a given phase of the oscillation) to the period of maximal receptiveness of the receiver area (the period of maximal excitability). One can intuit this as essentially analogous to the phenomenon of constructive interference, with in this case the interfering "waves" representing the locally generated oscillation and the input from an upstream oscillator. Conversely, oscillations which are out of phase or have no consistent phase relationship are less likely to coordinate output and input timing and therefore will facilitate less efficient information transfer. Said simply, coherence is a measure of how much two regions are "listening" to one another in a given frequency band [72].

We quantified the ratio of peak 4Hz coherence to peak 8Hz coherence as a measure of the relative likelihoods of these two coherent states over a given time period. We found, consistent with our own and others' previous results that fear conditioning increased the 4:8Hz coherence ratio (Figure 2.4 D) [74]. In other words, fear conditioning increased the 4Hz coherence relative to 8Hz, consistent with coherence in the 4Hz frequency band representing a state of fear. As expected, we found that extinction learning lead to a marked reduction in 4:8Hz coherence ratio, paralleling the reduction in fear behavior seen after extinction learning. Importantly, during exposure to the neutral context (CtxB), mice exhibited very low 4:8Hz coherence ratios (Figure 2.4 D). These results establish that the state of fear is characterized by high 4Hz coherence relative to 8Hz coherence, and that the relative balance of these coherent states is modified by

learning. One way to interpret these results is that they indicate the presence of a bistable or metastable multidimensional space, with 4Hz and 8Hz coherence representing the two dominant attractor states. In other words, the oscillatory activity across the mPFC-BLA circuit tends to settle, or "attract", in either a 4Hz or 8Hz pattern that is coherent across the circuit. From this perspective, the role of experience-dependent learning would be to modify the likelihood of settling in a given state as a result of a discrete set of sensory data. From an energetics perspective, this can be represented as modifiable peaks and troughs in an attractor landscape (Figure 2.5). I will elaborate on the utility of this abstraction further in the discussion, but I raise it now for its implications on our experimental interpretation. One would predict, for example, that the ability of 4Hz stimulation to shift the mPFC-BLA circuit towards a 4Hz coherent state would be modifiable, and in fact would likely rely on previous fear learning. Conversely, one would predict that the ability of 8Hz stimulation to shift mPFC-BLA coherence towards the 8Hz attractor would require extinction learning. Importantly, these predictions would parallel and perhaps partial explain our previous behavior findings, namely that 4 and 8Hz stimulation was capable of bidirectional fear behavior modulation only *after* previous fear and extinction learning, and only in the conditioned context. To test this, we quantified the effect of 4Hz and 8Hz stimulation (compared to no stimulation) on 4:8Hz coherence ratios in four behavioral states: prior to fear conditioning in the conditioning context (Pre-FC), during retrieval of the fear memory in the conditioned context (Post-FC), during retrieval following extinction learning (Retrieval A), and during exposure to a neutral context (Retrieval B). This way, we were able to quantify the effects on circuit activity of identical manipulation across multiple learning trials. If any difference in the

effect of stimulation emerged, it is reasonable to conclude that the difference is due to experience-dependent modification of the circuit -- in other words, learning. We found, in parallel to our behavioral data, that 4Hz stimulation during retrieval post-extinction learning was capable of inducing 4Hz coherence across the mPFC-BLA circuit. Similarly, 8Hz stimulation was capable of shifting induced 8Hz coherence (Figure 2.5 C,E). This is the first experimental demonstration that the 8Hz rhythm is capable of suppressing the 4Hz, pro-fear, state – something predicted from our previous data. In further support of this, we found that the ratio of 4Hz:8Hz coherence correlated with the amount of freezing in the retrieval trial (Figure 2.4 F). Strikingly, toggling this ratio by exogenous 4Hz or 8Hz stimulation shifted the freezing levels precisely as predicted. Even more strikingly, we found that identical manipulation in the neutral context B was not capable of robust, bidirectional shifts in coherence, nor did 4Hz:8Hz coherence ratio correlate with freezing levels (Figure 2.4 C,F-G). Finally, we found that 8Hz stimulation was not capable of inducing 8Hz coherence prior to extinction learning. Though a more comprehensive discussion of the implications of this will follow in the next section, suffice it to say that this demonstrates that learning fundamentally alters the capability of exogenous rhythmic manipulation to shift the coherence state of the mPFC-BLA circuit and thus affect behavioral output. This may give us insight into the mechanisms of experience-induced changes in circuit function that underlie learning.

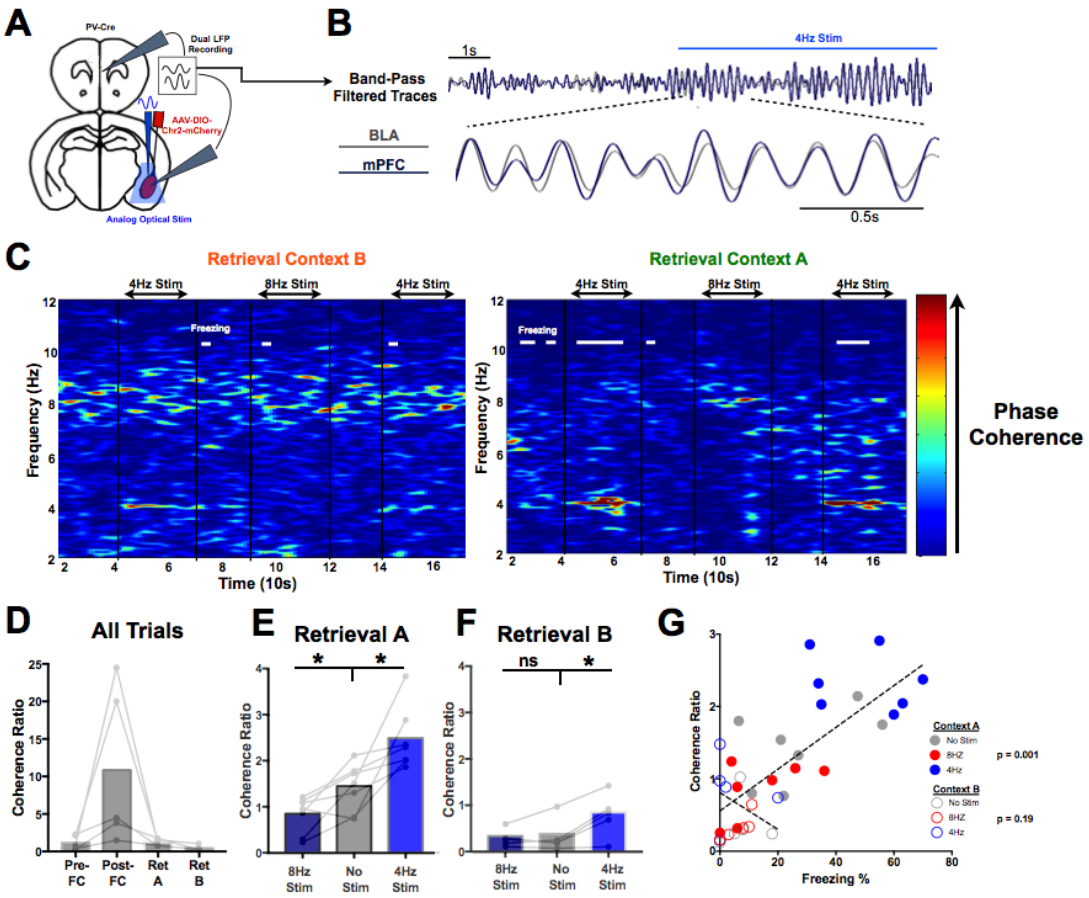

Figure 2.4. BLA PV-interneurons modulate mPFC-BLA coherence to control fear and extinction memory retrieval

A) Injection and recording strategy. AAV-DIO-Chr2-mCherry is injected unilaterally into the BLA of PV-IRES-Cre mice. Optical cannulas are placed into the BLA of injected animals, allowing frequency-specific analog stimulation of BLA PV-interneurons. Electrodes are placed in BLA and mPFC to allow simultaneous LFP recordings. **B)** Example overlaid band-pass (3-6Hz) filtered traces from BLA and mPFC before and during 4Hz analog optical stimulation. INSET: BLA and mPFC show high levels of coherence in the 4Hz range at the onset of 4Hz stimulation. **C)** Example phase coherograms from the same animal from retrieval in the conditioned context (A) and a neutral, unconditioned context (B), which demonsrate the ability to switch peak coherence states (heat map) and behavioral states (freezing indicated by white bars) by optical stimulation in context A, but not in context B. **D)** Fear conditioning increased the ratio of 4:8Hz coherence, while fear extinction decreases it. Animals exhibit consistently low 4:8Hz coherence ratios in context B. **E-G)** 4Hz stimulation increased 4:8Hz coherence ratio during retrieval in both context A and context B, while 8Hz stimulation decreased 4:8Hz coherence ratio only in context A. Importantly, 4:8Hz coherence ratio correlated tightly with freezing levels across animals in context A, but not context B. Retrieval A P = 0.04, 0.03. Retrieval B P = 0.81, 0.03. n = 7, 5. Paired t-tests, Dunnett's multiple comparison correction. Linear regression: Retrieval A $R^2 = 0.6$, P = 0.0001; Retrieval B $R^2 = 0.16$, P = 0.19

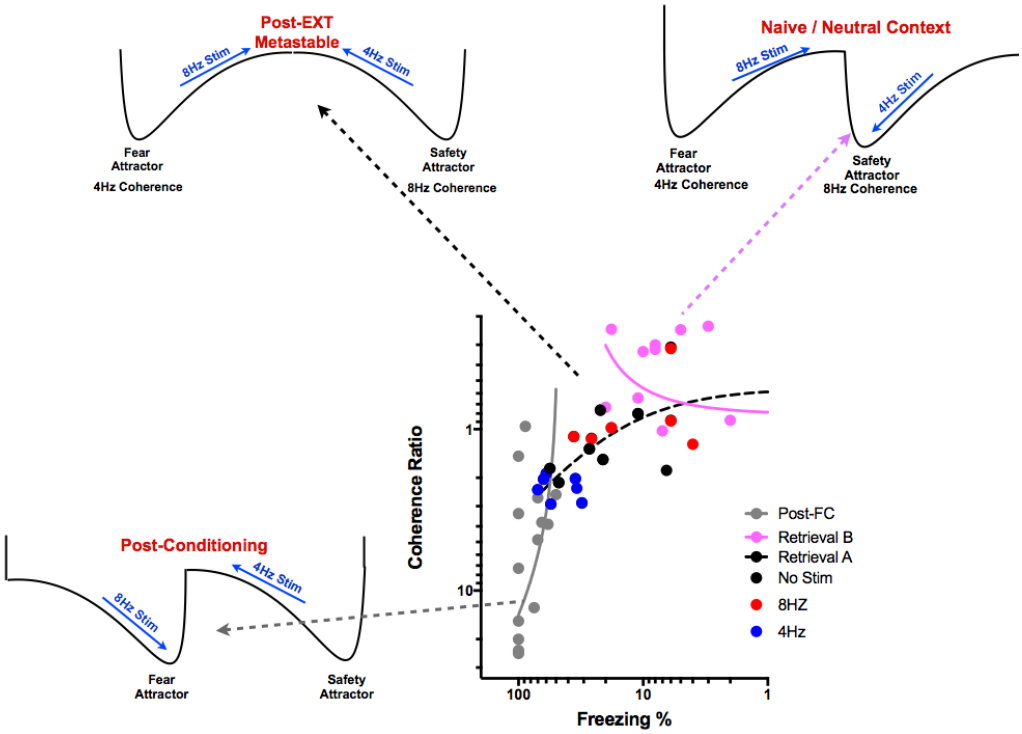

Figure 2.5. Extinction induces a metastable state across the mPFC-BLA circuit.
A) Plotting all data from Post-FC, Ret A, and Ret B together on a logarithmic scale allows visualization of the relationship between coherence state, stimulation, and behavior across multiple forms of learning. The cluster of pink dots represents states of low 4:8Hz coherence and low levels of freezing behavior. Fear conditioning causes a fundamental reorientation of this relationship such that a new cluster appears (gray dots) in states of high 4:8Hz coherence and high levels of freezing. Importantly, the slope of this relationship (gray line, linear regression) indicates the relationship between coherence and behavior. Since it is drastically different from Pre- and Post- FC conditions, we can conclude that fear conditioning alters the relationship between coherence states and behavior. The relative steepness of this relationship indicates that a) high coherence states are more likely than prior to FC, b) that multiple coherence states can produce the same behavior, and c) that even 8Hz stimulation of PV-interneurons, which reduces 4:8Hz coherence, does not alter behavior, indicating a relative robustness of the high fear, high 4:8Hz coherence state. This can be visualized as an attractor state in a two-dimensional field (bottom left), where the ability to alter the coherence state and therefore change behavior is represented by the slope of the line between the two attractor state (wells). Strikingly, extinction learning again alters the slope of this relationship (black dashed line), which again indicates a restructuring of the relationship between coherence states and behavior. Critically, only after extinction learning can 8Hz (red dots) or 4Hz (blue dots) bidirectionally shift coherence states to alter the behavioral state. Again, this can be visualized as the shape of the line between two attractors in a 2D field (top left, "Post-EXT Metastable"). The fact that the circuit can rapidly toggle between the two states, and that 4 or 8Hz stimulation can shift both coherence and behavior, can be represented as a less steep line between the two attractors. This ability to toggle rapidly and flexibly between states is characteristic of metastability, which is represented by the relatively flat "plateau" of the coherence vs. freezing curve in the post extinction state (black dashed line in graph, middle plateau in upper left schema).

Chapter 4: Discussion

Using a model of classical fear conditioning and extinction learning, and employing techniques focused across multiple levels of neural circuit function, we have uncovered new mechanistic insight into the encoding and modification of associative memories in the mPFC-BLA circuit. Specifically, our data support a model whereby extinction learning rearranges the structure of the BLA microcircuit, likely partially through inhibitory plasticity, to fundamentally alter the relationship between two functionally distinct network states and allow for the a "pro-safety" state to out-compete, or suppress, a "pro-fear" state. Further, these network states are characterized by their dominant frequency bands, and we provide direct evidence demonstrating the causal relationships of these states to behavioral output, as well as preliminary evidence linking the network state to the functional ensembles through the phenomenon of resonance. This model has several important implications and unresolved questions, some of which I will discuss below.

The principal implications of this work pertain to two critical questions: 1) how does extinction learning alter the BLA circuitry and 2) can this be understood across multiple "levels" in a way that enhances our understanding of the complex phenomena of neural circuits in general? I believe, based on our data in context with current understanding of neural circuit function, the answer to both of these lies interpreting our data through the lens of emergent circuit properties known as oscillation. Below I will briefly review the implications of this work on our understanding of neural oscillations and the relationships they have to functional ensembles and learning *per se*.

Inhibitory Plasticity and PV-interneurons

Previous work in our lab correlated target-specific structural changes in PV+ inhibitory synapses around fear-encoding neurons within the BLA in the suppression of these neurons and consequent suppression of fear behavior following extinction learning. The model put forward from the work implied that fear-encoding neurons within BLA recruit more inhibitory synapses during extinction learning, which causes them to be less active and therefore effect a reduced fear behavior state. This model is easy to intuit, as it is essentially a variation of the classical "excitatory/inhibitory balance" function of interneurons. In this view, interneurons act simply as a brake to excitatory behavior[65]. As we will see below, this view is inadequate and perhaps misleading. But first, let's draw attention to the significant contributions of this work.

Plasticity of excitatory synapses is well known due to the work of Eric Kandel and many others[3]. The most well known form of this plasticity is long term potentiation (LTP), whereby the post-synaptic potential produced by a given stimulus is increased, or potentiated, at specific synapses[93]. The work of Kandel and his descendants has shown, beyond a reasonable doubt, that sensory information can directly lead to changes in synaptic strength, which would presumably affect the input-output relationship of the circuit and could therefore be termed memory. In particular, the structural plasticity of excitatory synapses onto dendritic spines, which have been shown to be dynamic and a reasonable correlate of excitatory input, has been studied in great detail, in developing, healthy adult, and diseased brains[94]. However, a far less well studied, but presumably similarly important, subset of plasticity also exists at inhibitory synapses[95]. Because

inhibitory synapses do not have such easily identifiable morphological correlates of function as the spines of excitatory synapses, the structural changes that occur within them have only recently begun to be studied rigorously. Recent work has demonstrated structural plasticity of inhibitory synapses using fusion constructs to image inhibitory PSDs *in vivo* [96–98]. However, to date, the specific role of inhibitory plasticity in learning and memory has yet to be studied extensively. It is within this context that we should place the previous work in our lab, which was the first demonstration of experience-dependent inhibitory plasticity in the mammalian brain that could potentially underlie memory formation. Furthermore, it elegantly demonstrated the potential of inhibitory plasticity to affect specific functional ensembles of neurons, rather than the *en masse* dampening usually invoked. An intriguing line of investigation, then, would be to determine if exogenous induction of inhibitory plasticity in fear neurons with post-synaptic scaffolding proteins like neuroligin-2, for example, would lead to similar behavior and circuit level effects as extinction [99]. As mentioned above, however, the model of interneuronal suppression of fear-encoding neurons has several limitations. Primarily, it necessarily ignores the role of timing in inhibitory function. Indeed, the data outlined in this thesis, specifically correlating number of perisomatic PV+ synapses around BLA ensembles with their likelihood of activation (ZIF expression), demonstrate that the role of interneurons in controlling principal neuron activation cannot be explained merely by the number of synapses around its target neurons. As the example of rebound spiking demonstrates, inhibitory input onto principal neurons serves many functions, including synchronizing principal neuron spiking, grouping and segregating principal neuronal ensembles, and coordinating communication with distant circuits

through feedforward inhibition and oscillatory synchrony [58,65]. It is therefore necessary that we interpret our previous work, as well as the current work, in a manner that incorporates the diverse functions of interneurons in circuit computation. A more comprehensive and, as I will argue below, accurate, interpretation, is that BLA PV-interneurons are critical to the generation of oscillatory network states, and that plasticity of inhibitory synapses modifies the relationship between these network states in a manner that alters the functional output of BLA in an experience-dependent manner.

What do ensembles in BLA and mPFC encode?

One of the interesting questions addressed by our work and significant work of others is what, precisely, is encoded by the firing of BLA projection neurons? The canonical theory, as described in the introductory chapter, holds that BLA ensembles store the associative fear, or, in other cases, reward, memory [100]. This is accomplished by STDP converging on BLA principal neurons from CS and US carrying inputs. While there is no doubt that BLA plays a critical role in the formation of the associative memory, and that artificial activation of neurons can produce the behavioral effects interpreted as retrieval of the memory, this model has several limitations. First, although a full review of the topic is well beyond the scope of this discussion, it is well known that the "memory trace" – or at least the necessary components for retrieval of the memory – is dynamic and is in fact not stored in one location within the brain [101–106]. The mere passage of time, in addition to active updating mechanisms such as consolidation and reconsolidation, alter the necessity of individual components for memory retrieval. The best known example of this is hippocampal dependent memory, which requires proper

hippocampal activity during the learning and acute retrieval phases, but, given a long enough passage of time, can be retrieved in a hippocampus-independent manner [107]. This is interpreted as the memory "moving" or "distributing", usually to cortex. Second, the evidence from single unit data and other *in vivo* techniques indicates that BLA neurons fire in a rather complex fashion in response to external cues, which are not so easily delineated [100]. This problem is compounded by the fact that neurons clearly do not act as on-off switches in the brain, and rate encoding and other complex encoding possibilities are generally ignored in current models of BLA function. Finally, the brain cannot merely explicitly represent the environment, so some mechanism must be in place for internal representations of previous experience. So, then, if the memory is not simply "stored" in BLA, then what is it that BLA neurons encode, and, crucially – how? I certainly can't fully answer those questions, but I will argue that an alternative conception would be useful in interpreting the heretofore generated data.

Let's use the ILC TetTag neurons as an instructive example. As stated in the Results section, our data indicates that neurons in ILC are active in two states: during exposure to a novel context during which no negative experience is represented, and during exposure to context in which the animal has previously been fear conditioned, and therefore "expects" a negative experience, but does not experience one. These may be reconciled with the notion of stated-based prediction error (PE). PE is a well-known phenomenon in the reward literature, and represents a deviation from an expected experience or cue-experience relationship[108,109]. Stated-based error is distinct from canonical versions of prediction error, which typically emphasize reward signals and the difference between actual and expected reward. In reward prediction error theories, for

example, if a reward-predicting cue is presented without the associated reward, then there would be negative (reward) prediction error, or the unexpected absence of a predicted reward. A state prediction error may explain the activity of ILC ensembles by positing that the neurons are active when the current state differs from the expected state or state transition, without regard for the reward or "value" of the experience[110,111]. This would potentially enable flexible learning which does not depend only on reinforcement of existing behavioral strategies, because it depends on comparisons between experience and expected state-action-state transitions (paths). In addition, state-based learning models could potentially reconcile the activity of ILC neurons in a novel context (NS group) with their activity during the beginning of extinction learning, both of which could elicit strong state-based errors, as well as the known roles of ILC in both extinction learning and reward-based learning, both of which likely rely on interactions between an animal's expected experience and actual experience. Obviously, the experiments performed above were designed to investigate BLA ensemble encoding and behavior, and were not designed for detecting prediction errors, so they must therefore be interpreted cautiously. It will be interesting for follow-up work to investigate the potential for state-based learning to be encoded by ILC ensemble activity. Further studies, likely using single-unit recordings, are necessary to precisely delineate these necessary conditions for ILC ensemble engagement, as well as the role of BLA→mPFC projections therein. It is interesting to note that prediction-error-encoding cells have been reported in the amygdala, and the amygdala is also known to be involved in both reward and aversive conditioning [112 113]. Furthermore, the most convincing data on BLA ensemble function has been through manipulation of BLA principal neurons *during* learning, indicating that

BLA does not merely encode behavioral output, but likely conveys features critical to learning to other regions of the brain -- mPFC, for example -- possibly by comparing expected experience to current experience (eg: through state-based prediction errors) [36]. In any case, it is likely that the secret to mPFC-BLA circuit function lies in a more sophisticated understanding of these phenomena, an understanding that goes beyond artificially behaviorally-tethered abstractions. Furthermore, as I discuss below, a potential solution to these issues is to study, rather than just the on-off status of the ensembles themselves, the flexible and dynamic interactions between oscillations and the ensembles embedded within them.

Oscillations and Engrams

The work presented here clearly establishes the role of BLA PV-interneurons in generating oscillatory circuit phenomena, and that these oscillatory phenomena exert profound control over BLA output and learned behavior. We also provide evidence that, specifically after extinction learning, PV-interneurons exert selective control over functional subsets of pyramidal neurons within the BLA. These subsets, or "ensembles," are thought to be the basis of memory encoding within BLA and generally in cortical and hippocampal circuits [48] . Indeed, significant previous work has clearly established the existence of ensembles, or "engram cells", which appear to encode discrete features of the environment or even discrete memories, though this remains controversial. As discussed in the introduction and previous section, the most convincing evidence of the engram cells relies on sufficiency tests using artificial (chemo- or opto-genetic) activation of previously active, or IEG-expressing, subsets of neurons within a given brain region.

The seminal study in this blossoming field demonstrated that the behavioral effects of contextual fear conditioning could be elicited simply by optogenetically driving the activity of dentate gyrus neurons which expressed the IEG c-Fos during conditioning [54]. Much subsequent work, mostly from the same group, has pushed the model of engram cells as an explanation for the memory-encoding capabilities of cortex and hippocampus. From this view, a "memory" is merely the activity of a discrete set of principal neurons. This notion bears resemblance to the theory of "gnostic units", which held that individual and specialized neurons, or "units", encoded discrete memories and even percepts within the brain [114]. The abstraction is roughly similar to how a computer encodes information – in discrete physical locations within the computational circuit – thus the hypernym "engram". Memory, and therefore, learned behavior, then, is simply a matter of accessing the correct file. While this model is potentially useful for intuitive purposes there are limitations to this feedforward hierarchical abstraction: 1) It addresses the question of encoding by avoiding it. If a memory is encoded by the firing of a particular set of cells, then it is already "encoded" by the time those cells fire, and the memory will be accessed when those cells fire together. In other words, it leaves the question of "how does learning lead to the activity a functionally discrete subset of cells, and how does the firing of those cells bring about the internal or external manifestations of memory?" – in other words, the fundamental questions – for others to address. 2) Due to the technical limitations of using IEGs to label the target ensemble, it necessarily ignores the dynamic activation patterns that give rise to the formation and subsequent access of the engram. This is highlighted by the fact that the most common form of optogenetic reactivation is accomplished with square wave pulses at 20Hz, surely not a physiologic activation

pattern. 3) It does not explain the overlap or participation of single neurons in multiple representations, nor similar forms of flexibility, which are necessary for sufficient coding capacity. And 4) From an evolutionary perspective (taken cautiously, as all evolutionary arguments should be), it intuitively seems unlikely that the brain would evolve in such a way that limits coding capacity so clearly, and seems to imply that the evolutionary advantage of mammalian or even human brains is simply *more neurons,* which is unsupported structurally.

Which is all to say that something *like* engram cells clearly do exist, but likely not in the simplistic form usually posited. One way to potentially rectify this is to attempt to reconcile the emergent properties of the circuit -- oscillations -- with ensemble or engram theory. As usual, those who study the hippocampal circuit have advanced the field greatly, where many decades of work has demonstrated precise and complex relationships between CA1 place cells (engrams of the spatial variety) and the dominant oscillation of the hippocampus – the theta rhythms. By way of example, place cells fire in a particular sequence with clearly delineated phase relationships to theta (and other) cycles during an animal's exploration of an environment or learning of a spatial memory, and disruption of this relationship disrupts memory stability [49,115]. Regarding the relationship between oscillations and ensemble encoding in the amygdala, I believe the work presented here represents a significant contribution to our understanding. First, we demonstrate that the same manipulation – chemogenetic silencing of PV-interneurons – results in reactivation of a discrete subset of neurons, likely involved in the encoding and retrieval of the associative fear memory, as well as a shift in the oscillatory dynamics of the BLA towards a 4Hz, "pro-fear", oscillation. Interestingly, similar effects to both

ensemble activity and oscillatory structure were seen in mPFC, clearly indicating a connection between the oscillatory state and recruitment of "memory encoding" ensembles. Thankfully, we need not reinvent the wheel, and recent beautiful work has already shown that 4Hz oscillations in mPFC recruit single-units correlated with fear behavior – roughly analogous to our "FC-Tagged" population -- during the ascending phase of the rhythm [75] . Still, the question remains – how? Our work suggests that the answer may lie in part in phenomenon known as resonance. We provide evidence that distinct functional ensembles within BLA, or at least neurons active during different and opposing forms of learning, have distinct intrinsic resonance profiles. One would predict that, among other critical parameters, the intrinsic resonance profile, or tuning curve, of a neuron would bias its inclusion into an ongoing rhythm. For example, neurons tuned to 4Hz input would be more likely to be recruited during that oscillation and therefore participate in fear behavior. Conversely, neurons tuned to 8Hz input might participate in a distinct rhythmogenic circuit with opposing output consequences. Although these data are preliminary, we provide some evidence that indeed this may be exactly the case – that one can predict the recruitment of BLA ensembles into oscillations based on their intrinsic tuning curves. This, in addition to other critical parameters – the amount of perisomatic inhibition each neuron receives, for example – may explain why the 4Hz and 8Hz oscillations have opposing behavioral consequences. This notion is appealing for several reasons. First, it potentially reconciles much previous literature regarding oscillations and ensembles in the BLA and mPFC. For example, it could potentially assist in explaining the findings of the Josselyn group that exogenous CREB expression biases inclusion into subsequent memory ensembles, and from the Silva group that the

overlap of CA1 ensembles encoding discrete events decreases as a function of time [115–117]. These could both be explained if exogenous CREB expression (former) or learning induced-Fos expression (latter) modifies the intrinsic properties of the neurons to bias their inclusion into learning-relevant oscillations going forward. Interesting, this very notion was proposed many years ago, but has apparently been washed away by the LTP tidal wave [118]. Still, it is intriguing to note that recent work from the Josselyn group has demonstrated the PV-interneuron activity can also alter inclusion into a memory trace, perhaps linking the circuit and "intrinsic" aspects of ensemble formation [119]. Second, it may explain how "competition" between the two states is possible, by positing that fear-encoding and extinction-encoding neurons, together with PV-interneurons (and likely many other cell types not discussed here), form a lateral- and feedback-inhibitory circuit [61,65] . A feedback loop with 4Hz-tuned cells would compete for the same shared interneuron population with a distinct population of 8Hz-tuned cells. The "winner", or dominant rhythm, would emerge through the optimal combination of excitatory drive (likely with frequency information itself), intrinsic resonance, inhibitory synaptic input, and spike timing relative to excitation of the other principal neuron populations. If this sounds familiar, that is likely because it invokes the mechanisms behind "winner-take-all" gamma oscillations and rebound spiking, both of which depend critically on fast-spiking interneurons [69,120]. And finally, pursuant the previous point, such a microcircuit structure would allow multiple parameters, including alteration of intrinsic resonance by channel modification, and amount of perisomatic inhibition as seen in the previous work from our lab, to modify the relationship between the two oscillatory states. Such modification of behaviorally meaningful network states would be the essence of learning.

Future work will need to confirm these preliminary findings, address the local structures at play, and investigate more fully what the consequences of learning are on these parameters.

Non-linearities and resonant engrams

Part of the appeal of the converging models of STDP/LTP basis of memory, put forth by the pioneering work of Eric Kandel, and engram (ensemble) theory, as epitomized by the recent work from the Tonegawa group, is that they are easy to intuit separately and also in collaboration. It is easy to see how these two could be merged to satisfy the original postulates of Donald Hebb's "cell assemblies" theory, which holds that encoding in the brain is achieved by activity of transient partnerships among neurons [6]. These partnerships are brought about from modifications of synaptic strength during learning and development to increase the likelihood that the participant members would subsequently "fire together". From this view, the role of STDP during learning is to create ensemble engrams, which then will be re-activated (retrieved) more easily during subsequent exposure to the inciting set of sensory stimuli (ie: the learned experience). An alternative, though not mutually exclusive, proposition, is that STDP alters the oscillatory properties of the network in question, and the likelihood of ensemble participation in a given oscillatory state. From this perspective, STDP does ultimately achieve the Hebbian goal of a transient assembly, but not necessarily through direct synaptic connection – rather through the participant neurons' relationships with ongoing oscillations. By way of example – in the BLA, neurons which are biased, in an experience-dependent manner, through STDP and/or alterations in intrinsic properties, towards inclusion into a given

oscillation, will then fire with distinct temporal relationships with one another via the rhythm. If these temporal relationships are on the short time scale of a gamma rhythm, for example, then they will be indistinguishable to downstream neurons due to the temporal scales of dendritic integration. Even longer rhythms, like the 4Hz and 8Hz states described here, can recruit ensembles at given phases of the oscillation, as shown by the previously mentioned work from the Herry lab, or by the common feature of neuronal networks known as phase-amplitude coupling – that is, the amplitude of a higher frequency oscillation can be coupled to the phase of a lower-frequency oscillation [75,121,122]. In this way, even slower rhythms can recruit ensemble activity that is, as far as the post-synaptic neuron is concerned, simultaneous. Oscillations can achieve synchrony of neuronal spiking, a key element of circuit function, without necessitating a "synfire chain" of participating neurons. If the participants are recruited to the same oscillation, they need not themselves be directly connected[123].

Which is all to say that the relationships of individual neurons to oscillatory states, and the relationships between the states themselves, represents an important avenue of investigation regarding circuit function. The work presented here contributes to this investigation in several ways. First, our findings indicate that the return of fear following extinction learning is characterized by a few notable distinctions from the outwardly similar fear behavior following conditioning. Though both fear states are characterized by high 3-6Hz activity and low 6-12Hz activity, the return of fear following extinction is marked by a discrete signal from BLA to mPFC in the 3-6Hz band. This indicates that, although the "attractor state" of 3-6Hz oscillatory activity and fear behavior is similar, the "path" towards that state is fundamentally altered by extinction

learning. In the pre-extinction state, mPFC appears to entrain BLA in that frequency band, a notion supported by the fact that optogenetic induction of 4Hz activity in dmPFC leads to freezing behavior irrespective of previous conditioning, but equivalent induction in BLA does not[58]. In the post-extinction state, however, BLA appears to entrain mPFC in the same frequency band. In both cases, the circuit converges (attracts) to a 3-6Hz oscillatory state, but their "paths" are distinct. In other words, the mPFC-BLA circuitry is structurally capable of facilitating both 3-6Hz pro-fear and 6-12Hz pro-safety oscillations at baseline, but learning modulates the likelihood of existing in one of these attractor states by modulating the paths taken between them. This would be consistent with our previous finding that only after extinction learning do the two states exhibit a negative correlation or "competition" with one another – a property evidently dependent on local PV-interneurons. In other words, extinction learning fundamentally alters the relationship between the two attractor states. In addition the general notion that extinction learning alters the interaction or "paths" between these two attractor states would also be consistent with our previous findings that BLA PV-interneurons are required to mediate fear suppression following, but not before, extinction learning. Finally, this would potentially explain why optogenetic manipulation of BLA PV-interneurons produces robust bidirectional behavioral effects only after, but not before, extinction learning. From this perspective, the same optogenetic manipulation would not necessarily produce the same behavioral effects before and after learning because the relationship between the attractor states has been altered. This may be why, for example, 8Hz stimulation is only capable of suppressing 4Hz activity and freezing behavior *after* extinction learning, and is only capable of doing so in the conditioned context. Said in another way, the substrate

of learning is in the modification of the interactions and paths between pre-existing attractor states.

To make this case more intuitively, let's imagine the two attractor states as energy wells, similar to those abstracted from thermodynamic principles (Figure 2.6). The depth of the well would represent its stability – that is, robustness to both external perturbation and stochastic fluctuation. Similarly, the shape of the path (line) between two states, addition to the height of the intervening peak, would represent the available paths (trajectories) taken between the two states. In thermodynamics, this is manifest in the likelihood of a given reaction taking place, which requires the system to move from one energetic trough to another. In our case, it is manifest in the likelihood of a state transition from one attractor (fear, 4Hz coherence) to another (safety, 8Hz coherence). In its essence, the shape and height of the transition trajectories are what we are probing by exogenous 4Hz or 8Hz manipulation. Put simply, the fact that the same exact manipulation has little to no effect before learning, and a robust effect following learning, indicates a modified trajectory between these two states. This validates the interpretation that learning is represented in experience-dependent changes in state transition trajectories (or paths) between attractor states. The continuation of this interpretation, then, is that the post-conditioning state is characterized by an energetically stable fear attractor state with very limited trajectories towards the safety state (Figure 2.6, bottom left). This would be consistent with the fact that the animal stays in a state of fear for prolonged periods of times (stability) following conditioning, and with our observation that the power of the 3-6Hz and 6-12Hz bands are not correlated following conditioning. This lack of correlation would be because minor fluctuations in the power of the 6-12Hz

oscillation are not sufficient to transition the circuit to the safety attractor state. This is known as robustness. Conversely, the post-extinction state is characterized by a negative correlation between the powers of the two oscillations, indicating that fluctuations in the power of one oscillation one are frequently capable of inducing a state-transition. This is known as flexibility, and is supported by our data indicating that the animals which exhibited the most negative correlation between 3-6Hz and 6-12Hz powers following extinction learning were the ones that still exhibited some (>10%) freezing (data not shown). A completely stable and robust safety attractor would exhibit no consistent correlation, because minor fluctuations in 3-6Hz power would not be able to affect the 6-12Hz power sufficiently to effect a state transition. From this perspective, extinction learning induces a circuit state which is easily able to transition from one attractor to another (Figure 2.6, top left). In non-linear dynamics, this is known as "metastability"[124,125]. Metastability is characterized by an inability to "settle" into one of the available attractor states, which distinguishes it from bistability or multistability. This feature allows rapid switching between output states, as well as rapid combining and modification of output structures. A metastable state would exhibit anti-correlation between the markers of the attractors, and would be comparatively easy to force transitions within via exogenous manipulation. Our data support the notion that extinction learning induces a metastable state between the two pre-existing attractor states of fear and safety. Our comparative advantage in identifying this is the clearly delineated internal (4Hz vs 8Hz) and external (freezing) behaviors characteristic of these states, and which in turn allowed this analysis. However, there is no reason in principle to exclude the interpretation that similar phenomena underlie learning and memory of a much more

complex and adaptive variety. Indeed, metastability has already been proposed as a hallmark feature of complex computational circuits in the neocortex[124,125]. Further work will be necessary to determine whether the findings outlined here will be relevant to such cortical output behaviors as semantic memory, decision-making, and cognition.

While these certainly sound like lofty abstractions, there are several reasons not to disregard them as such. First, the "engram" theory is equivalently abstract and in fact relies less on experimental evidence and endogenous activity patterns based in primary observation. Second, whether we choose to ignore the fact or not, the brain is a complex (non-linear) device, and a reasonable argument can be made that, given the near impossibility of bottom-up understanding of such complex systems (due to the difficult to predict relationship of input and output), a more direct strategy is to experimentally address the emergent phenomena of those systems and built theoretical models based principally on them. Finally, on a practical level, we have experimental access to the rhythms of human brains much more so than we have access to its synaptic and cellular substituents. In addition, technologies are emerging that allow experimental manipulation of said rhythms, which has already proven useful for certain pathologies[126–128]. From a translational perspective, exogenous induction of rhythmic activity during, for example, exposure therapy, seems an exciting therapeutic avenue for the treatment of anxiety disorders, PTSD, and other fear pathologies. Additionally, our models predicts that interruption of BLA→mPFC communication in the 3-6Hz range may prevent the spontaneous return of fear following exposure therapy. Intuitively, this seems more immediately achievable than the cell-type specific optogenetic strategies usually proposed in NIH grants. In addition, oscillatory activity represents a natural

endophenotype for the screening and design of pharmacologic interventions. For these and more reasons, anyone who is seriously interested in "bench-to-bedside" fear circuitry research should give thought to oscillatory phenomena. A genuine understanding of their properties, relationships to circuit output and ultimate behavioral production, as well as how the lower-level components give rise to, and are affected by, oscillatory activity, would be a boon for basic and translational neuroscientists alike. It is my sincere hope that the work summarized here contributes ever-so-slightly to that endeavor.

Chapter 5: Bibliography

1. Schacter, D. L. Understanding implicit memory. A cognitive neuroscience approach. *Am Psychol* **47,** 559–69 (1992).
2. SCOVILLE, W. B. & MILNER, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatr.* **20,** 11–21 (1957).
3. Kandel, E. R., Dudai, Y. & Mayford, M. R. The molecular and systems biology of memory. *Cell* **157,** 163–86 (2014).
4. Mackintosh, N. J. Pavlov and associationism. *Span J Psychol* **6,** 177–84 (2003).
5. Dickinson, A. & Mackintosh, N. J. Classical conditioning in animals. *Annu Rev Psychol* **29,** 587–612 (1978).
6. Hebb, D. The organization of behavior: A neuropsychological theory. (2005). at <http://books.google.com/books?hl=en&lr=&id=uyV5AgAAQBAJ&oi=fnd&pg=PP1&ots=mIuUElFUWm&sig=0UeSMfjnpRByyY2Gj0zu43zPG6Q>
7. Martin, S. J., Grimwood, P. D. & Morris, R. G. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23,** 649–711 (2000).
8. Bailey, C. H., Kandel, E. R. & Harris, K. M. Structural Components of Synaptic Plasticity and Memory Consolidation. *Cold Spring Harb Perspect Biol* **7,** a021758 (2015).
9. Lashley, KS. The problem of serial order in behavior. *Cerebral mechanisms in behavior* (1951). at <http://pubman.mpdl.mpg.de/pubman/item/escidoc:2352626/component/escidoc:2352625/Lashley_1951_The_Problem_of_Serial_Order_in_Behavior.pdf>
10. Hawkins, R. D., Kandel, E. R. & Bailey, C. H. Molecular mechanisms of memory storage in Aplysia. *Biol. Bull.* **210,** 174–91 (2006).
11. Lomo, T. Potentiation of monosynaptic EPSPs in the perforant path-dentate granule cell synapse. *Exp Brain Res* **12,** 46–63 (1971).
12. Bliss, T. V. & Lomo, T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol. (Lond.)* **232,** 331–56 (1973).
13. Nicoll, R. A. & Malenka, R. C. Contrasting properties of two forms of long-term potentiation in the hippocampus. *Nature* **377,** 115–8 (1995).
14. Shapiro, M. Plasticity, hippocampal place cells, and cognitive maps. *Arch. Neurol.* **58,** 874–81 (2001).
15. Nakazawa, K., McHugh, T. J., Wilson, M. A. & Tonegawa, S. NMDA receptors, place cells and hippocampal spatial memory. *Nat. Rev. Neurosci.* **5,** 361–72 (2004).
16. Mayford, M., Siegelbaum, S. A. & Kandel, E. R. Synapses and memory storage. *Cold Spring Harb Perspect Biol* **4,** (2012).
17. Pittenger, C. & Kandel, E. A genetic switch for long-term memory. *C. R. Acad. Sci. III, Sci. Vie* **321,** 91–6 (1998).
18. LeDoux, J. The emotional brain, fear, and the amygdala. *Cell. Mol. Neurobiol.* **23,** 727–38 (2003).

19. Parkes, S. L. & Westbrook, R. F. Role of the basolateral amygdala and NMDA receptors in higher-order conditioned fear. *Rev Neurosci* **22,** 317–33 (2011).
20. Goosens, K. A. & Maren, S. NMDA receptors are essential for the acquisition, but not expression, of conditional fear and associative spike firing in the lateral amygdala. *Eur. J. Neurosci.* **20,** 537–48 (2004).
21. Maren, S., Aharonov, G., Stote, D. L. & Fanselow, M. S. N-methyl-D-aspartate receptors in the basolateral amygdala are required for both acquisition and expression of conditional fear in rats. *Behav. Neurosci.* **110,** 1365–74 (1996).
22. LaLumiere, R. T., McGaugh, J. L. & McIntyre, C. K. Emotional Modulation of Learning and Memory: Pharmacological Implications. *Pharmacol. Rev.* **69,** 236–255 (2017).
23. Morris, R. G. NMDA receptors and memory encoding. *Neuropharmacology* **74,** 32–40 (2013).
24. Asede, D., Bosch, D., Lüthi, A., Ferraguti, F. & Ehrlich, I. Sensory inputs to intercalated cells provide fear-learning modulated inhibition to the basolateral amygdala. *Neuron* **86,** 541–54 (2015).
25. Papp, E. *et al.* Glutamatergic input from specific sources influences the nucleus accumbens-ventral pallidum information flow. *Brain Struct Funct* **217,** 37–48 (2012).
26. Bienvenu, T. C., Busti, D., Magill, P. J., Ferraguti, F. & Capogna, M. Cell-type-specific recruitment of amygdala interneurons to hippocampal theta rhythm and noxious stimuli in vivo. *Neuron* **74,** 1059–74 (2012).
27. Gore, F. *et al.* Neural Representations of Unconditioned Stimuli in Basolateral Amygdala Mediate Innate and Learned Responses. *Cell* **162,** 134–45 (2015).
28. Bienvenu, T. C. *et al.* Large intercalated neurons of amygdala relay noxious sensory information. *J. Neurosci.* **35,** 2044–57 (2015).
29. Wolff, S. B. *et al.* Amygdala interneuron subtypes control fear learning through disinhibition. *Nature* **509,** 453–8 (2014).
30. Nabavi, S. *et al.* Engineering a memory with LTD and LTP. *Nature* **511,** 348–52 (2014).
31. Janak, P. H. & Tye, K. M. From circuits to behaviour in the amygdala. *Nature* **517,** 284–92 (2015).
32. Felix-Ortiz, A. C. *et al.* BLA to vHPC inputs modulate anxiety-related behaviors. *Neuron* **79,** 658–64 (2013).
33. Namburi, P. *et al.* A circuit mechanism for differentiating positive and negative associations. *Nature* **520,** 675–8 (2015).
34. Kim, J., Zhang, X., Muralidhar, S., LeBlanc, S. A. & Tonegawa, S. Basolateral to Central Amygdala Neural Circuits for Appetitive Behaviors. *Neuron* **93,** 1464–1479.e5 (2017).
35. Kim, J., Pignatelli, M., Xu, S., Itohara, S. & Tonegawa, S. Antagonistic negative and positive neurons of the basolateral amygdala. *Nat. Neurosci.* **19,** 1636–1646 (2016).
36. Senn, V. *et al.* Long-range connectivity defines behavioral specificity of amygdala neurons. *Neuron* **81,** 428–37 (2014).
37. Gvozdanovic, G. A., Stämpfli, P., Seifritz, E. & Rasch, B. Neural correlates of experimental trauma memory retrieval. *Hum Brain Mapp* (2017). doi:10.1002/hbm.23613

38. Esber, G. R. & Holland, P. C. The basolateral amygdala is necessary for negative prediction errors to enhance cue salience, but not to produce conditioned inhibition. *Eur. J. Neurosci.* **40,** 3328–37 (2014).
39. Watanabe, N., Sakagami, M. & Haruno, M. Reward prediction error signal enhanced by striatum-amygdala interaction explains the acceleration of probabilistic reward learning by emotion. *J. Neurosci.* **33,** 4487–93 (2013).
40. Furlong, T. M., Cole, S., Hamlin, A. S. & McNally, G. P. The role of prefrontal cortex in predictive fear learning. *Behav. Neurosci.* **124,** 574–86 (2010).
41. Orsini, C. A. & Maren, S. Neural and cellular mechanisms of fear and extinction memory formation. *Neurosci Biobehav Rev* **36,** 1773–802 (2012).
42. Woo, H. J., Yu, C., Kumar, K. & Reifman, J. Large-scale interaction effects reveal missing heritability in schizophrenia, bipolar disorder and posttraumatic stress disorder. *Transl Psychiatry* **7,** e1089 (2017).
43. Milad, M. R. & Quirk, G. J. Fear extinction as a model for translational neuroscience: ten years of progress. *Annu Rev Psychol* **63,** 129–51 (2012).
44. Quirk, G. J. *et al.* Erasing fear memories with extinction training. *J. Neurosci.* **30,** 14993–7 (2010).
45. Goode, T. D. & Maren, S. Animal models of fear relapse. *ILAR J* **55,** 246–58 (2014).
46. LeDoux, J. The amygdala. *Curr. Biol.* **17,** R868–74 (2007).
47. Tovote, P., Fadok, J. P. & Lüthi, A. Neuronal circuits for fear and anxiety. *Nat. Rev. Neurosci.* **16,** 317–31 (2015).
48. Tonegawa, S., Liu, X., Ramirez, S. & Redondo, R. Memory Engram Cells Have Come of Age. *Neuron* **87,** 918–31 (2015).
49. Moser, M.-B. B., Rowland, D. C. & Moser, E. I. Place cells, grid cells, and memory. *Cold Spring Harb Perspect Biol* **7,** a021808 (2015).
50. HUBEL, D. H. & WIESEL, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol. (Lond.)* **148,** 574–91 (1959).
51. Reijmers, L. G., Perkins, B. L., Matsuo, N. & Mayford, M. Localization of a stable neural correlate of associative memory. *Science* **317,** 1230–3 (2007).
52. Farivar, R., Zangenehpour, S. & Chaudhuri, A. Cellular-resolution activity mapping of the brain using immediate-early gene expression. *Front. Biosci.* **9,** 104–9 (2004).
53. Mayford, M. & Reijmers, L. Exploring Memory Representations with Activity-Based Genetics. *Cold Spring Harb Perspect Biol* **8,** a021832 (2015).
54. Liu, X. *et al.* Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **484,** 381–5 (2012).
55. Cho, J.-H. H., Deisseroth, K. & Bolshakov, V. Y. Synaptic encoding of fear extinction in mPFC-amygdala circuits. *Neuron* **80,** 1491–507 (2013).
56. Pinard, C. R., Mascagni, F. & McDonald, A. J. Medial prefrontal cortical innervation of the intercalated nuclear region of the amygdala. *Neuroscience* **205,** 112–24 (2012).
57. Trouche, S., Sasaki, J. M., Tu, T. & Reijmers, L. G. Fear extinction causes target-specific remodeling of perisomatic inhibitory synapses. *Neuron* **80,** 1054–65 (2013).
58. Hu, H., Gan, J. & Jonas, P. Interneurons. Fast-spiking, parvalbumin^{+} GABAergic interneurons: from cellular design to microcircuit function. *Science* **345,** 1255263 (2014).

59. Hu, H. & Jonas, P. A supercritical density of Na(+) channels ensures fast signaling in GABAergic interneuron axons. *Nat. Neurosci.* **17,** 686–93 (2014).
60. Woodruff, A. R. & Sah, P. Networks of parvalbumin-positive interneurons in the basolateral amygdala. *J. Neurosci.* **27,** 553–63 (2007).
61. Smith, Y., Paré, J. F. & Paré, D. Differential innervation of parvalbumin-immunoreactive interneurons of the basolateral amygdaloid complex by cortical and intrinsic inputs. *J. Comp. Neurol.* **416,** 496–508 (2000).
62. Ryan, S. J. *et al.* Spike-timing precision and neuronal synchrony are enhanced by an interaction between synaptic inhibition and membrane oscillations in the amygdala. *PLoS ONE* **7,** e35320 (2012).
63. Woodruff, A. R. & Sah, P. Inhibition and synchronization of basal amygdala principal neuron spiking by parvalbumin-positive interneurons. *J. Neurophysiol.* **98,** 2956–61 (2007).
64. Stark, E. *et al.* Inhibition-induced theta resonance in cortical circuits. *Neuron* **80,** 1263–76 (2013).
65. Roux, L. & Buzsáki, G. Tasks for inhibitory interneurons in intact brain circuits. *Neuropharmacology* **88,** 10–23 (2015).
66. Cobb, W. A. Evolution of clinical neurophysiology since Hans Berger, the past 40 years of EEG. *Electroencephalogr Clin Neurophysiol* **27,** 648–9 (1969).
67. Cohen, M. X. Where Does EEG Come From and What Does It Mean? *Trends Neurosci.* **40,** 208–218 (2017).
68. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13,** 407–20 (2012).
69. Buzsaki, G. Rhythms of the Brain. (2006). at <http://books.google.com/books?hl=en&lr=&id=ldz58irprjYC&oi=fnd&pg=PA4&dq=Rhythms+of+the+Brain+Buzsaki&ots=Q1-646hNDR&sig=FNSOOY3jvUILoSXyxDORLsf7I4c>
70. Klausberger, T. & Somogyi, P. Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* **321,** 53–7 (2008).
71. Fries, P. Rhythms for Cognition: Communication through Coherence. *Neuron* **88,** 220–35 (2015).
72. Harris, A. Z. & Gordon, J. A. Long-range neural synchrony in behavior. *Annu. Rev. Neurosci.* **38,** 171–94 (2015).
73. Teleńczuk, B. *et al.* Local field potentials primarily reflect inhibitory neuron activity in human and monkey cortex. *Sci Rep* **7,** 40211 (2017).
74. Karalis, N. *et al.* 4-Hz oscillations synchronize prefrontal-amygdala circuits during fear behavior. *Nat. Neurosci.* **19,** 605–12 (2016).
75. Dejean, C. *et al.* Prefrontal neuronal assemblies temporally control fear behaviour. *Nature* **535,** 420–4 (2016).
76. Lesting, J. *et al.* Directional theta coherence in prefrontal cortical to amygdalo-hippocampal pathways signals fear extinction. *PLoS ONE* **8,** e77707 (2013).
77. Lesting, J. *et al.* Patterns of coupled theta activity in amygdala-hippocampal-prefrontal cortical circuits during fear extinction. *PLoS ONE* **6,** e21714 (2011).
78. Seidenbecher, T., Laxmi, T. R., Stork, O. & Pape, H.-C. C. Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval. *Science* **301,** 846–50 (2003).

79. Stujenske, J. M., Likhtik, E., Topiwala, M. A. & Gordon, J. A. Fear and safety engage competing patterns of theta-gamma coupling in the basolateral amygdala. *Neuron* **83,** 919–33 (2014).
80. Likhtik, E., Stujenske, J. M., Topiwala, M. A., Harris, A. Z. & Gordon, J. A. Prefrontal entrainment of amygdala activity signals safety in learned fear and innate anxiety. *Nat. Neurosci.* **17,** 106–13 (2014).
81. Brunner, C., Delorme, A. & Makeig, S. Eeglab - an Open Source Matlab Toolbox for Electrophysiological Research. *Biomed Tech (Berl)* (2013). doi:10.1515/bmt-2013-4182
82. Armbruster, B., Li, X., Pausch, M., Herlitze, S. & Roth, B. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc National Acad Sci* **104,** 5163–5168 (2007).
83. Ehrlich, D. E., Ryan, S. J. & Rainnie, D. G. Postnatal development of electrophysiological properties of principal neurons in the rat basolateral amygdala. *J. Physiol. (Lond.)* **590,** 4819–38 (2012).
84. McGarry, L. M. & Carter, A. G. Inhibitory Gating of Basolateral Amygdala Inputs to the Prefrontal Cortex. *J. Neurosci.* **36,** 9391–406 (2016).
85. Wall, N. R., Wickersham, I. R., Cetin, A., De La Parra, M. & Callaway, E. M. Monosynaptic circuit tracing in vivo through Cre-dependent targeting and complementation of modified rabies virus. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 21848–53 (2010).
86. Giustino, T. F. & Maren, S. The Role of the Medial Prefrontal Cortex in the Conditioning and Extinction of Fear. *Front Behav Neurosci* **9,** 298 (2015).
87. Adhikari, A. *et al.* Basomedial amygdala mediates top-down control of anxiety and fear. *Nature* **527,** 179–85 (2015).
88. Laurent, V. & Westbrook, R. F. Inactivation of the infralimbic but not the prelimbic cortex impairs consolidation and retrieval of fear extinction. *Learn. Mem.* **16,** 520–9 (2009).
89. Do-Monte, F. H., Manzano-Nieves, G., Quiñones-Laracuente, K., Ramos-Medina, L. & Quirk, G. J. Revisiting the role of infralimbic cortex in fear extinction with optogenetics. *J. Neurosci.* **35,** 3607–15 (2015).
90. Thompson, B. M. *et al.* Activation of the infralimbic cortex in a fear context enhances extinction learning. *Learn. Mem.* **17,** 591–9 (2010).
91. Milad, M. R. & Quirk, G. J. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* **420,** 70–4 (2002).
92. Adhikari, A., Sigurdsson, T., Topiwala, M. A. & Gordon, J. A. Cross-correlation of instantaneous amplitudes of field potential oscillations: a straightforward method to estimate the directionality and lag between brain areas. *J. Neurosci. Methods* **191,** 191–200 (2010).
93. Siegelbaum, S. A. & Kandel, E. R. Learning-related synaptic plasticity: LTP and LTD. *Curr. Opin. Neurobiol.* **1,** 113–20 (1991).
94. Sala, C. & Segal, M. Dendritic spines: the locus of structural and functional plasticity. *Physiol. Rev.* **94,** 141–88 (2014).
95. Frias, C. P. P. & Wierenga, C. J. Activity-dependent adaptations in inhibitory axons. *Front Cell Neurosci* **7,** 219 (2013).

96. Chen, J. L. *et al.* Clustered dynamics of inhibitory synapses and dendritic spines in the adult neocortex. *Neuron* **74,** 361–73 (2012).
97. Chen, J. L. & Nedivi, E. Highly specific structural plasticity of inhibitory circuits in the adult neocortex. *Neuroscientist* **19,** 384–93 (2013).
98. Van Versendaal, D. *et al.* Elimination of inhibitory synapses is a major component of adult ocular dominance plasticity. *Neuron* **74,** 374–83 (2012).
99. Poulopoulos, A. *et al.* Neuroligin 2 drives postsynaptic assembly at perisomatic inhibitory synapses through gephyrin and collybistin. *Neuron* **63,** 628–42 (2009).
100. Gründemann, J. & Lüthi, A. Ensemble coding in amygdala circuits for associative learning. *Curr. Opin. Neurobiol.* **35,** 200–6 (2015).
101. Wang, S.-H. H. & Morris, R. G. Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annu Rev Psychol* **61,** 49–79, C1–4 (2010).
102. Bergstrom, H. C. The neurocircuitry of remote cued fear memory. *Neurosci Biobehav Rev* **71,** 409–417 (2016).
103. Dash, P. K., Hebert, A. E. & Runyan, J. D. A unified theory for systems and cellular memory consolidation. *Brain Res. Brain Res. Rev.* **45,** 30–7 (2004).
104. Paré, D. Role of the basolateral amygdala in memory consolidation. *Prog. Neurobiol.* **70,** 409–20 (2003).
105. Fries, P., Fernández, G. & Jensen, O. When neurons form memories. *Trends Neurosci.* **26,** 123–4 (2003).
106. Tulving, E. & Markowitsch, H. J. Memory beyond the hippocampus. *Curr. Opin. Neurobiol.* **7,** 209–16 (1997).
107. Squire, L. R. & Bayley, P. J. The neuroscience of remote memory. *Curr. Opin. Neurobiol.* **17,** 185–96 (2007).
108. Schultz, W. Dopamine reward prediction error coding. *Dialogues Clin Neurosci* **18,** 23–32 (2016).
109. Chase, H. W., Kumar, P., Eickhoff, S. B. & Dombrovski, A. Y. Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cogn Affect Behav Neurosci* **15,** 435–59 (2015).
110. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66,** 585–95 (2010).
111. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69,** 1204–15 (2011).
112. McHugh, S. B. *et al.* Aversive prediction error signals in the amygdala. *J. Neurosci.* **34,** 9024–33 (2014).
113. McNally, G. P., Johansen, J. P. & Blair, H. T. Placing prediction into the fear circuit. *Trends Neurosci.* **34,** 283–92 (2011).
114. Barlow, H. B. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* **38,** 795–8 (2009).
115. Sosa, M., Gillespie, A. K. & Frank, L. M. Neural Activity Patterns Underlying Spatial Coding in the Hippocampus. *Curr Top Behav Neurosci* (2016). doi:10.1007/7854_2016_462
116. Yiu, A. P. *et al.* Neurons are recruited to a memory trace based on relative neuronal excitability immediately before training. *Neuron* **83,** 722–35 (2014).

117. Han, J.-H. H. *et al.* Selective erasure of a fear memory. *Science* **323,** 1492–6 (2009).
118. Cai, D. J. *et al.* A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* **534,** 115–8 (2016).
119. Destexhe, A. & Marder, E. Plasticity in single neuron and circuit computations. *Nature* **431,** 789–95 (2004).
120. Rashid, A. J. *et al.* Competition between engrams influences fear memory formation and recall. *Science* **353,** 383–7 (2016).
121. De Almeida, L., Idiart, M. & Lisman, J. E. A second function of gamma frequency oscillations: an E%-max winner-take-all mechanism selects which cells fire. *J. Neurosci.* **29,** 7497–503 (2009).
122. Canolty, R. T. & Knight, R. T. The functional role of cross-frequency coupling. *Trends Cogn. Sci. (Regul. Ed.)* **14,** 506–15 (2010).
123. Tort, A. B., Komorowski, R., Eichenbaum, H. & Kopell, N. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *J. Neurophysiol.* **104,** 1195–210 (2010).
124. Buzsáki, G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* **68,** 362–85 (2010).
125. Deco, G. & Kringelbach, M. L. Metastability and Coherence: Extending the Communication through Coherence Hypothesis Using A Whole-Brain Computational Perspective. *Trends Neurosci.* **39,** 125–35 (2016).
126. Tognoli, E. & Kelso, J. A. The metastable brain. *Neuron* **81,** 35–48 (2014).
127. Fang, J. Y. & Tolleson, C. The role of deep brain stimulation in Parkinson's disease: an overview and update on new developments. *Neuropsychiatr Dis Treat* **13,** 723–732 (2017).
128. Miller, S. & Matharu, M. Non-invasive Neuromodulation in Primary Headaches. *Curr Pain Headache Rep* **21,** 14 (2017).
129. Dalkilic, E. B. Neurostimulation Devices Used in Treatment of Epilepsy. *Curr Treat Options Neurol* **19,** 7 (2017).