Received: 7 September 2023 | Revised: 3 February 2024 | Accepted: 11 March 2024

DOI: 10.1111/ejn.16338

RESEARCH REPORT

EJN European Journal of Neuroscience FENS WILEY

A scalable spiking amygdala model that explains fear conditioning, extinction, renewal and generalization

Peter Duggins[1,2] | **Chris Eliasmith**[1,2,3]

[1]Centre for Theoretical Neuroscience, University of Waterloo, Waterloo, Ontario, Canada

[2]Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada

[3]Department of Philosophy, University of Waterloo, Waterloo, Ontario, Canada

Correspondence
Peter Duggins, Centre for Theoretical Neuroscience, University of Waterloo, Waterloo, Ontario, Canada.
Email: psipeter@gmail.com

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: 261453; Canadian Foundation for Innovation, Grant/Award Number: 52479-10006; Air Force Office of Scientific Research, Grant/Award Number: FA9550-17-1-0026; Ontario Innovation Trust, Grant/Award Number: 35768

Edited by: Bernard Balleine

Abstract

The amygdala (AMY) is widely implicated in fear learning and fear behaviour, but it remains unclear how the many biological components present within AMY interact to achieve these abilities. Building on previous work, we hypothesize that individual AMY nuclei represent different quantities and that fear conditioning arises from error-driven learning on the synapses between AMY nuclei. We present a computational model of AMY that (a) recreates the divisions and connections between AMY nuclei and their constituent pyramidal and inhibitory neurons; (b) accommodates scalable high-dimensional representations of external stimuli; (c) learns to associate complex stimuli with the presence (or absence) of an aversive stimulus; (d) preserves feature information when mapping inputs to salience estimates, such that these estimates generalize to similar stimuli; and (e) induces a diverse profile of neural responses within each nucleus. Our model predicts (1) defensive responses and neural activities in several experimental conditions, (2) the consequence of artificially ablating particular nuclei and (3) the tendency to generalize defensive responses to novel stimuli. We test these predictions by comparing model outputs to neural and behavioural data from animals and humans. Despite the relative simplicity of our model, we find significant overlap between simulated and empirical data, which supports our claim that the model captures many of the neural mechanisms that support fear conditioning. We conclude by comparing our model to other computational models and by characterizing the theoretical relationship between pattern separation and fear generalization in healthy versus anxious individuals.

KEYWORDS
anxiety disorder, empirical validation, neural engineering framework, spiking neuron model

Abbreviations: -, extinction; *, novel; +, acquisition; AMY, amygdala; BLA, basolateral amygdala; BNST, bed nucleus of the stria terminalis; CeL, central lateral amygdala; CeM, central medial amygdala; CS, conditioned stimulus; CTX, context; EMG, electromyography; err, error neurons; EXT, external populations; fMRI, functional magnetic resonance imaging; GAD, generalized anxiety disorder; IL, infralimbic cortex; inh, inhibitory neurons; ITC, intercalated cells; LA, lateral amygdala; LIF, leaky-integrate-and-fire; mPFC, medial prefrontal cortex; MTL, medial temporal lobe; NEF, neural engineering framework; PES, prescribed error sensitivity; pyr, pyramidal neurons; UR, unconditioned response; US, unconditioned stimulus.

1 | INTRODUCTION

Fear conditioning is a process where animals learn to associate a neutral stimulus with an aversive (unpleasant) stimulus, such that they exhibit a defensive response when presented with the neutral stimulus. It is one of the most common forms of learning in the animal kingdom and has been widely studied in neuroscience and psychology. The amygdala (AMY) is an important locus of fear conditioning in the brain: It undergoes synaptic plasticity during conditioning, and lesions to it impair both learning and expression of conditioned responses. Furthermore, AMY modulates cognitive processes like perception, attention and decision-making by orchestrating the brain's emotional and hormonal responses to salient stimuli. Understanding learning within AMY is therefore central to the development of integrated theories of affective cognition and the treatment of anxiety disorders.

In this paper, we develop a neural model of fear conditioning in AMY that includes numerous biological details, suggests explainable representations and learning processes and makes several neural and behavioural predictions. We hypothesize that fear learning within AMY emerges as a result of error-driven learning on the synapses between specific AMY nuclei, and we propose a computational model of the required neural mechanisms to test this theory. Anatomically, our network contains populations of excitatory and inhibitory neurons, organized into distinct nuclei, whose internal and external connectivity are appropriately constrained; we show that these neurons naturally develop empirically consistent response profiles over the course of simulated experiments. Functionally, our network learns to associate aversive events with complex stimuli and environmental contexts through the use of a novel, scalable learning rule that accommodates stimuli containing multiple features. After training our model using standard conditioning protocols, we test the network's degree of fear acquisition, extinction and renewal. From these data, we generate predictions in three distinct domains. First, we predict defensive responses to various stimuli and environmental contexts. Second, we predict that ablating specific nuclei (pharmacologically or electromagnetically inactivating their neurons) causes specific deficits in fear learning and expression. Third, we predict the extent to which an individual will generalize a defensive response to similar stimuli and characterize how the shape of this gradient depends on the pattern separation induced by neural tuning curves. Specifically, our model predicts that when neurons outside AMY are overtuned to a wide range of external stimuli, AMY will exhibit weak pattern separation to potential stressors, resulting in the overgeneralized defensive responses observed in anxiety disorders.

We begin by reviewing the literature on fear conditioning and summarizing the anatomical properties of AMY, focusing in particular on the features required for our functional model. We then introduce the model, describe our simulated training protocols and run a variety of computational experiments. From these simulated results, we derive several predictions and then test those predictions by comparing simulated data to various classes of empirical data. We find that our predictions are well supported by the existing data and propose future experiments when such data are unavailable. We conclude by comparing our model with other neural models of fear conditioning; discussing the novelty of our approach with respect to biological realism, explainability and scalability; acknowledging its limitations; and identifying directions for future work.

2 | BACKGROUND

2.1 | Fear conditioning, extinction and generalization

Fear conditioning begins by placing an animal into some environment or 'context'. A neutral stimulus, referred to as the 'conditioned stimulus' (CS), is presented for a short duration, and the animal's baseline defensive response is recorded. During the 'aquisition' or 'conditioning' phase, the CS is presented alongside an aversive stimulus, referred to as the 'unconditioned stimulus' (US); the duration, ordering and overlap of the CS and US are experimental parameters. After multiple 'pairings' of the CS and US, the animal learns that the CS predicts the onset of the US and begins to exhibit the defensive response to the CS itself. During a 'fear expression' test, the CS is presented alone (without the US), and the animal's defensive response is again measured; if the response is greater than during the baseline test, fear conditioning has occurred.

Conditioned responses to the CS persist in long-term memory and are difficult to unlearn. In 'extinction' training, an animal is fear conditioned in the acquisition context (CTX^+), then undergoes a second round of training in some other context. The features of this extinction context are an important experimental parameter: The extinction context may be identical to the acquisition context (CTX^+), or it may be a new environment with readily distinguishable features (CTX^-). The CS is repeatedly presented without the US, and over time, the animal's defensive response to the CS diminishes. However, neural and behavioural evidence suggest that extinction

does not unlearn the original CS-US association[2]: Rather, the animal learns a separate 'safety' association that competes with the fear association and suppresses the defensive response (Milad & Quirk, 2012; Myers & Davis, 2002). This suppression is context specific: When the animal is tested in the extinction context, defensive responses are suppressed, but when tested in any other context, defensive responses return. This phenomenon is called 'fear renewal'.

Finally, 'fear generalization' (Webler et al., 2021) is the tendency for animals to exhibit defensive responses to stimuli (environments) that resemble the CS (CTX^+). For instance, animals conditioned to associate a large circle (CS^+) with the US, but a small circle (CS^-) with no US, will exhibit graded defensive responses to circles of intermediate sizes (Dunsmoor & Paz, 2015; Lissek et al., 2014). These results suggest that fear learning acts on a sub-symbolic level of abstraction: Stimuli that share features with the CS^+ (CTX^+) elicit partial defensive responses.

2.2 | Amygdala neuroanatomy

In this section, we describe how the internal and external connectivity of neural populations within AMY, as well as the interplay between excitatory pyramidal neurons and inhibitory interneurons, are critical for fear conditioning (Krabbe et al., 2018; Lucas & Clem, 2018; McDonald, 2020). While numerous other brain areas, such as medial temporal lobe (MTL), the infralimbic and prelimb cortices (IL and PL), the bed nucleus of the stria terminalis (BNST) and the intercalated cells, are implicated in fear conditioning and coordinating an animal's overall defensive response (Luchkina & Bolshakov, 2019), our computational model focuses on the intrinsic connectivity within AMY that supports fear learning and simple defensive behaviours. In Section 5.2, we discuss in detail other aspects of neuroanatomy that are important for fear conditioning and should be considered in future work.

AMY contains several principal nuclei, including the lateral amygdala (LA), basolateral amygdala (BLA), central lateral amygdala (CeL) and central medial amygdala (CeM), as shown in Figure 1. While other subdivisions of AMY are possible (e.g., by cell type McDonald, 2020), these four nuclei are the most frequently identified in theoretical accounts of fear conditioning. LA receives connections from sensory cortices that convey information about the external world (McDonald, 2020). The synapses between long-range cortical axons and the dendrites of pyramidal cells in LA are the primary site of acquisition; plasticity within these synapses induces long-term potentiation in response to coincident CS and US, increasing the sensitivity of (some) LA neurons to the CS and driving downstream defensive responses (Krabbe et al., 2018; Muller et al., 1997; Ressler & Maren, 2019). LA pyramidal neurons project to BLA and CeL but do not synapse with neurons outside AMY (Duvarci & Pare, 2014).

BLA receives fewer connections from sensory cortices but is reciprocally connected to several structures in the MTL (including the hippocampus) (McDonald, 2020) and mPFC (including IL and PL) (Mattera et al., 2020). These connections are thought to convey high-level contextual information to the BLA, including episodic memory from hippocampus and personal identifiers from MTL, and to facilitate consolidation, allowing the BLA to coordinate and control defensive responses based on context (Bocchio et al., 2017; Carrere & Alexandre, 2015; Duvarci & Pare, 2014; Vlachos et al., 2011). In particular, BLA is the primary site of contextual extinction: Synaptic plasticity on BLA pyramidal neurons and/or inhibitory interneurons is associated with decreased defensive responses in safe contexts. Pyramidal neurons within BLA exhibit several characteristic responses: 'Fear neurons' become more responsive to CS during acquisition but become less responsive during extinction; 'extinction neurons' do not respond to the CS following acquisition but become responsive during extinction; and 'persistent neurons' become CS-responsive during acquisition but remain responsive after extinction (Duvarci & Pare, 2014; McDonald, 2020). Unlike LA, BLA neurons project to several brain areas outside AMY: In addition to reciprocal connections with MTL, BLA outputs also reach striatum, which then projects back to cortex, suggesting triangular connections between BLA, cortex and striatum (McDonald, 2020). Through these connections, BLA is implicated in affective labelling, goal-directed behaviour, planning and decision-making (Mirolli et al., 2010). BLA also connects to CeM and CeL (directly and via interneurons) (Duvarci & Pare, 2014).

CeL receives external projections from thalamus and brainstem and internal projections from the LA and ITC. Like LA and BLA, the responsiveness of CeL neurons changes during acquisition, indicating that fear conditioning also induces synaptic plasticity within this nucleus (Keifer Jr et al., 2015). Unlike LA and BLA, CeL contains exclusively inhibitory interneurons whose reciprocal connections control defensive responses (Carrere & Alexandre, 2015; Keifer Jr et al., 2015). Baseline activity within CeL 'off' neurons inhibits CeM, preventing a default defensive response, but inactivating these 'off'

[2]However, unlearning may occur before long-term memory consolidation (Myers et al., 2006); we discuss this further in Section 5.1.

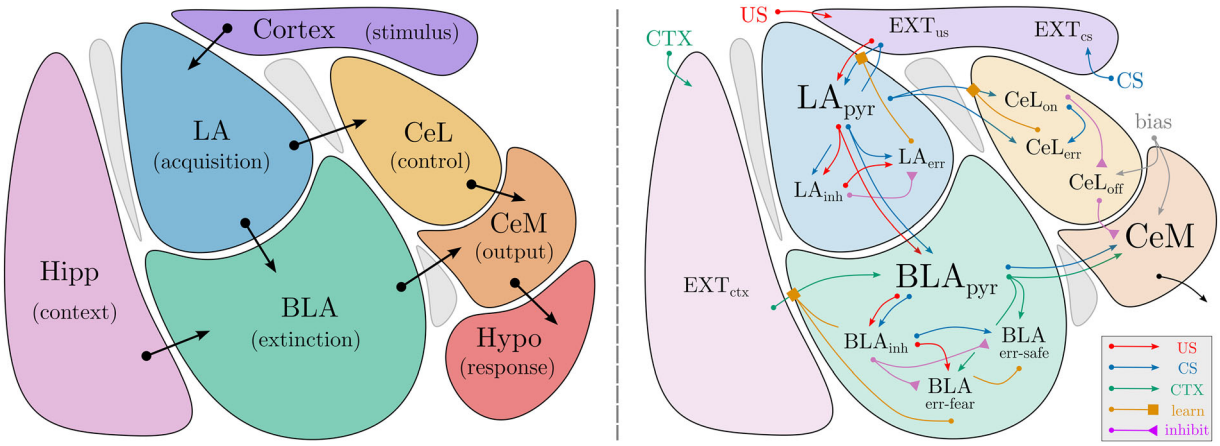

FIGURE 1 Left: Sketch of amygdala neuroanatomy. Nuclei abbreviations: LA, lateral amygdala; BLA, basolateral amygdala; CeL, central lateral amygdala; CeM, central medial amygdala. External population abbreviations: Hipp, hippocampus, medial temporal lobe and medial prefrontal cortex; Cortex, sensory cortices; Hypo, hypothalamus, periaqueductal gray, sympathetic and parasympathetic nervous system. The grey regions are intercalated cells (ITC). Right: Anatomically detailed model of fear conditioning in the amygdala. Black text indicates neural populations. Coloured arrows indicate the communication of information or the application of a particular function. See text for details about the representations used in each population and the functions computed by each connection.

neurons (or activating 'on' neurons that inhibit the 'off' neurons) produces an unconditioned defensive response (Ciocchi et al., 2010). CeL may thus act as a controllable gate for CeM-driven defensive responses.

CeM is the major output hub of AMY: It receives connections from BLA and CeL that convey fear and safety signals and projects to numerous subcortical structures that drive basic defensive responses. These targets include the hypothalamus, which controls endocrine responses in the brain and body; the periaqueductal gray, which directs behavioural responses to internal and external stressors; and numerous other structures in the sympathetic and parasympathetic nervous system (McDonald, 2020). Through the CeM, the AMY is able to regulate the body's state, to modulate large portions of the brain and to direct behavioural responses (Mirolli et al., 2010). In fear conditioning experiments, CeM appears to integrate the responses learned in LA, BLA and CeL to coordinate many aspects of an animal's subsequent defensive response (Duvarci & Pare, 2014). Consistent with this theory, activating CeM before training produces unconditioned freezing, while inactivating it after training reduces CS-induced freezing (Ciocchi et al., 2010).

2.3 | Amygdala function

Broadly speaking, AMY is part of the brain's 'affective system': It helps organisms decide which of their goals to pursue in the current context. This is likely achieved by using associative learning to assign 'salience' to objects, people and events in the wider world. Organisms routinely experience stimuli that have intrinsic or extrinsic value: Appetitive stimuli like food and sex are naturally desirable, while aversive stimuli like pain are undesirable. These stimuli should be approached, avoided or otherwise acted upon to increase an organism's chances of survival and reproduction. We expand on a functional hypothesis put forward by other theorists and experimentalists, which states that the AMY assigns salience to neutral stimuli by learning to associate their presence with these 'charged' stimuli then triggering intrinsic behavioural responses (Duvarci & Pare, 2014; John et al., 2016; Mirolli et al., 2010). In fear conditioning, AMY learns to trigger the animal's innate defensive response (or 'unconditioned' response [UR]) when the CS is reliably paired with the charged US. CS-UR associative learning guarantees that an organism will rapidly deploy responses to salient stimuli, independent of environmental context. This learning is realized in central amygdala (CeA): Connections from brainstem and thalamus to CeL convey simple representations of US and CS, while outputs from CeM to hormonal systems direct the UR (Keifer Jr et al., 2015).

As cortex evolved to use more sophisticated representations of the external world, we believe AMY evolved more sophisticated mechanisms of assigning salience and controlling behaviour via LA and BLA. High-dimensional representations in sensory cortex are conveyed to LA (McDonald, 2020), which learns to associate these CS (including auditory, olfactory and visual signals) with various types of US (Duvarci & Pare, 2014); heightened responses in LA then drive CeM (via CeL or BLA) to

activate the UR. Compared with CS-UR association in CeA, this pathway uses more complex representations of the CS, allowing an organism to distinguish stimuli that predict the US from those that are present by happenstance.

A third form of associative learning occurs in BLA, which helps AMY further distinguish salient features of the environment from irrelevant distractors. BLA receives sophisticated representations about an organism's context (including episodic and emotional memories) from MTL and mPFC (Bocchio et al., 2017; McDonald, 2020). However, it is difficult to directly account for this contextual information when learning CS-US associations: We suspect that learning three-way associations between CTX, CS and US is more complicated than learning a series of two-way relations between CTX/CS and CS/US. More importantly, incorporating CTX information directly into CS-US learning rules may cause an organism to incorrectly identify situations as non-salient. Neural and behavioural evidence suggest that BLA learns separate CTX associations that suppress defensive responses (Duvarci & Pare, 2014; Milad & Quirk, 2012; Myers & Davis, 2002): LA associations between CS and US are preserved during extinction training, but BLA learns a second memory trace that uses CTX information to correct false positives. This approach ensures that if the CTX association turns out to be wrong, or if more complicated contingencies arise, the organism does not have to relearn the original CS-US association; it simply amends the secondary memory trace in BLA. Indeed, the diverse response profiles of BLA neurons may support contextually-sophisticated salience assessments, possibly implementing a simple form of model-based control over fear expression (Prévost et al., 2013). Cortical areas like MTL and mPFC, whose hierarchical neural architectures and high-dimensional representations are well suited for model-based reasoning, may also support complex associations between CS, US and CTX and relay associative information to AMY through the BLA. Finally, contextual learning in AMY may be supported by reciprocal connections to, and consolidation within, IL and PL (Mattera et al., 2020; Pendyam et al., 2013; Sierra-Mercado et al., 2011; Sotres-Bayon & Quirk, 2010)

Taken together, this hypothesis prescribes the roles of the various AMY nuclei in an evolutionarily plausible manner and shows how parallel learning rules may coordinate diverse responses to salient stimuli. This account is consistent with other models of AMY function, both conceptual (Mirolli et al., 2010) and computational (Carrere & Alexandre, 2015; Duvarci & Pare, 2014). In the next section, we propose a computational model that realizes this account and tests a more specific hypothesis: that AMY uses error-driven learning rules to estimate the salience of external inputs using a particular representational scheme. The goal of this model is therefore to rigorously specify the functional capabilities described qualitatively in Section 2.3 while adhering to the anatomical constraints described in Section 2.2. Given that these two objectives are sometimes at odds (i.e., adding more biological detail may not improve function and may hinder explainability), we focus on building a functional model of AMY itself: Each population in our spiking neuron model has a well-defined functional role and is biologically constrained, but we do not describe the populations and connections outside AMY in detail, nor do we include biological components within AMY whose functional purpose we cannot surmise. We construct our model using theoretical tools that facilitate scalability and modularity, allowing future research to expand on the functionality and biological plausibility of the current work: We discuss the particulars of these extensions in Sections 5.1 and 5.2.

3 | MATERIALS AND METHODS

3.1 | Neural engineering framework (NEF)

Our AMY model is built using the NEF (Eliasmith & Anderson, 2003), a highly successful theory for building biologically constrained neural networks with functional capabilities (Eliasmith, 2013). Previous NEF models have been used to study reinforcement learning in cortex and basal ganglia (Duggins et al., 2022; Rasmussen, 2014), associative learning in cortex (Borst et al., 2018; Stewart et al., 2011; Voelker et al., 2014), temporal and spatial learning in hippocampus (Dumont et al., 2023, 2022), spike-timing-dependent plasticity (Bekolay et al., 2013; MacNeil & Eliasmith, 2011), adaptive motor control (DeWolf, 2015) and end-to-end learning of complex cognitive tasks in the world's largest functional brain model (Eliasmith et al., 2012). NEF models facilitate explanability by (a) grounding symbol-like representations in spiking neural activity and (b) grounding learning in the synaptic connections between populations of neurons: As a result, the relationship between the cognitive capacity of a model, and its biological implementation in a neural network, is always well defined. Finally, NEF models frequently provide testable predictions that have been validated in subsequent empirical experiments; for instance, an NEF model of path integration (Conklin & Eliasmith, 2005) has been credited with predicting the existence of grid cells (Zilli, 2012) and the result that path integration is independent of head direction (Maurer et al., 2014). Together, these features make NEF models well suited to the study of learning in the biological brain.

The NEF characterizes spiking activity within populations of neurons as encoding information in a latent *state space*. While spikes are the physical means of communication between neurons, cognition can be analysed as transformations of these states, permitting a more compact description of what brains do (Boerlin et al., 2013; Gallego et al., 2017; Recanatesi et al., 2022). We assume that states in this state space can be represented by a vector-valued signal $\mathbf{x}(t)$ and that the cognitive operations performed in the brain may be described as dynamical transformations of $\mathbf{x}(t)$.

The NEF begins by defining methods for encoding and decoding between neural activity and the state space. A neuron will fire most frequently when presented with its particular 'preferred stimulus' and will respond less strongly to increasingly dissimilar stimuli. Each simulated neuron i is accordingly assigned a preferred direction vector or *encoder*. When driven with an external signal $\mathbf{x}(t)$, the firing rate of the neuron is given by

$$a_i(t) = G[\alpha_i \mathbf{e}_i \cdot \mathbf{x}(t) + \beta_i], \tag{1}$$

where $a_i(t)$ is the spiking activity of neuron i, G is the neuron model with electrical current inputs $[\cdot], \alpha_i$ is the gain, β_i is the bias current and $\mathbf{e}_i \cdot \mathbf{x}(t)$ is the dot product between the state space inputs and neuron's encoder. A distributed encoding extends the notion of representation: If $\mathbf{x}(t)$ is fed into multiple neurons, each with a unique tuning curve defined by $\mathbf{e}_i, \alpha_i$ and β_i, then each neuron will respond with a unique spiking pattern $a_i(t)$, and the collection of all neural activities will robustly encode the signal.

In order to recover, or decode, the state space information encoded in neural spike trains, the NEF also defines *decoders* $\mathbf{d}_i$, which either perform this recovery or compute arbitrary functions, $f(\mathbf{x})$, of the represented vector. A functional decoding with $\mathbf{d}_i^f$ allows networks of neurons to *transform* the signal into a new state, which is essential for performing cognitive operations. To compute these transformations, a linear decoding is applied to the neural activities:

$$\hat{f}(\mathbf{x}(t)) = \sum_{i=0}^{n} a_i(t) \mathbf{d}_i^f, \tag{2}$$

where $a_i(t)$ is the activity of neuron i, n is the number of neurons and the hat notation indicates that the computed quantity is an estimate of the target function. Connection *weights* between each presynaptic neuron i and each postsynaptic neuron j are composed of encoders and decoders:

$$\mathbf{w} = \mathbf{e} \times \mathbf{d}^f. \tag{3}$$

In our AMY model, encoders are randomly distributed across a unit hypersphere with dimensionality D, ensuring that the neural population will effectively represent a state space signal $\mathbf{x} \in \mathbb{R}^D$. We use two methods to choose decoders. For connections that compute well-defined functions, we perform an offline least-squares optimization to solve for decoders that minimize the error between the neural estimate and the target function (see Eliasmith & Anderson, 2003 for details). For connections that learn associations, we use an online, spike-based, error-driven learning rule called the prescribed error sensitivity (PES) rule (MacNeil & Eliasmith, 2011):

$$\Delta\omega_{ij} = \Delta\mathbf{d}_i \cdot \alpha_j \mathbf{e}_j = \frac{\epsilon}{n} \alpha_j \mathbf{e}_j \cdot a_i(\hat{\mathbf{x}}(t) - \mathbf{x}(t)), \tag{4}$$

where ϵ is the decoder learning rate, n is the number of presynaptic neurons, a_i is the filtered activity of the presynaptic neuron $i, \mathbf{x}(t)$ is the state space target and $\hat{\mathbf{x}}(t)$ is the estimate decoded with Equation (2). Conceptually, Equation (4) uses the error between the decoded estimate and state space target to update the decoders during the simulation. Extensive work with the PES rule in NEF networks has shown it is capable of learning decoders to compute a wide range of functions (Voelker, 2015), and the PES rule has been shown to reproduce neural and behavioural data in brain areas ranging from cortex (Duggins et al., 2022) to hippocampus (Dumont et al., 2022) to cerebellum (Stöckel et al., 2021). As we discuss in detail in Section 5.1, the application of a scalable error-driven learning rule to fear conditioning in AMY is a novel contribution of this work: While most other computational model of AMY use Hebbian rules to learn associations between simple stimuli (see Section 5.3), the PES rule allows our model to learn complex associations between feature-rich stimuli. As a result, our model makes novel predictions about fear generalization in AMY that are outside the scope of previous models.

3.2 | Network structure and function

Our AMY model aims to realize the functionality described in Section 2.3 while respecting the anatomy described in Section 2.2. The complete network architecture is shown in Figure 1 (right panel). In this section, we introduce and analyse the components in a piecewise manner to clarify how they act in isolation.

Input signals conveying CS, US and CTX information drive three external (EXT) populations. The CS signal has dimensionality $D = 3$ and is presented for a duration of 1 s, followed by 1 s of silence (all zeros). The US signal is one dimensional (indicating the presence or absence of

the US) and is presented alongside the CS during acquisition (1 s on, then 1 s off). The CTX signal has dimensionality $D = 5$ and remains constant during acquisition, then changes during extinction or testing. Note that these D are free parameters of the model that represent the number of features present in the CS and CTX signals: Higher dimensional signals imply more sophisticated hierarchies for sensory processing that produce more complex sensory (or temporal) representations, but scaling D does not change the structure of the model. We found that three to five features per signal was sufficiently complex to perform the generalization experiments in Section 4.6. To generate the vectors corresponding to CS and CTX, we drew samples from the surfaces of D-dimensional hyperspheres: This ensures that the magnitude of the population response induced by different stimuli is approximately equivalent, even though each stimulus activates a different subset of EXT neurons. The dynamics of stimulus presentation are shown in Figure 2 (first panel).

We begin by learning CS-US associations in the LA. LA contains three populations: LA_{pyr} is excitatory pyramidal neurons that (may) become CS-sensitive during fear acquisition, LA_{inh} is inhibitory interneurons that gate learning and LA_{err} is neurons that calculate the required update signal. All three populations receive information about the CS and US, either directly (from EXT_{cs} or EXT_{us}) or indirectly (from LA_{pyr} or LA_{inh}). LA_{pyr} and LA_{inh} represent two-dimensional vectors (CS and US response), while LA_{err} represents a one-dimensional vector (CS response error). The US is

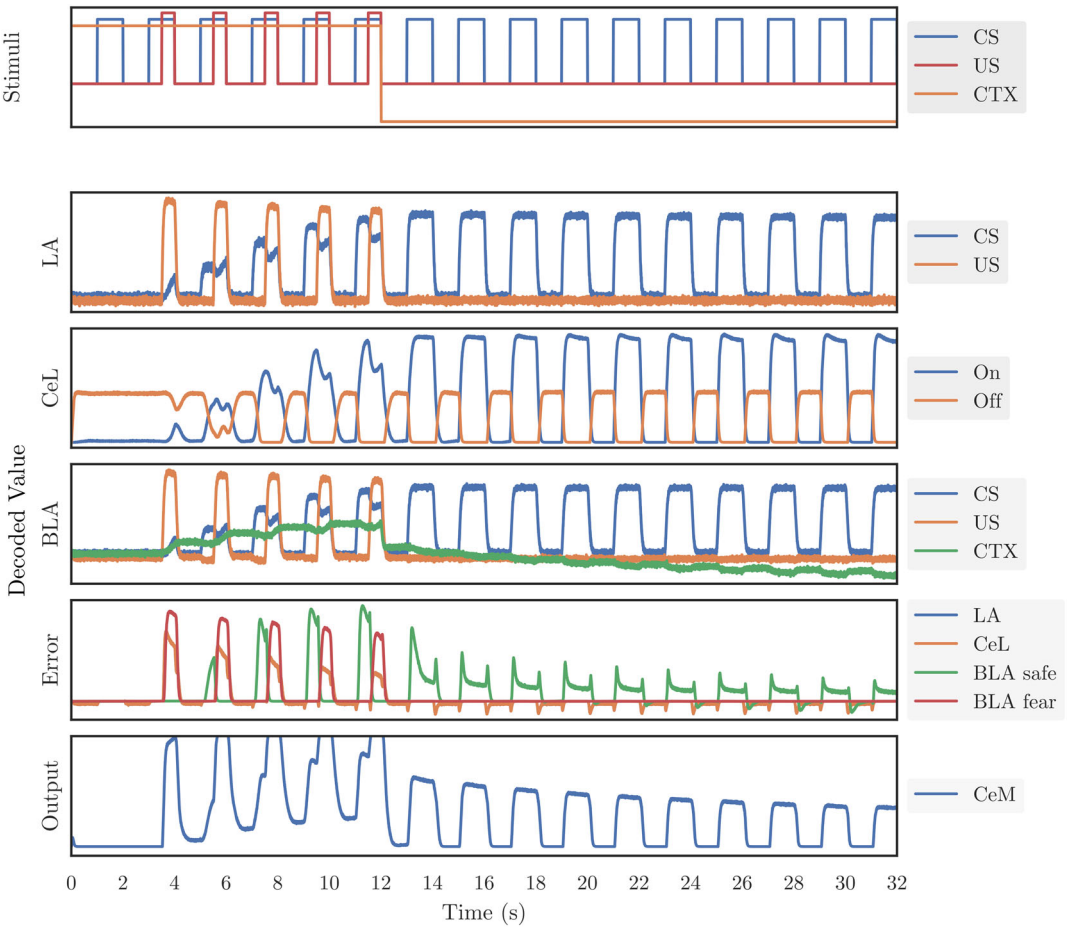

FIGURE 2 Dynamics of fear acquisition and extinction. The first panel indicates when CS, US and CTX are applied. Plotted CS values indicate the presence of the three-dimensional CS signal, while plotted CTX values are the dot product between the current context and CTX^{+}, both of which are five-dimensional vectors. The remaining panels plot $\hat{\mathbf{x}}(t)$, the values decoded from the spiking activity of the indicated populations (Equation 2). These unitless estimates can be used to decipher the internal representations in the model. During acquisition, PES learning causes LA_{pyr}'s CS response to rise until the value represented in LA_{err} reaches zero, and these responses persist in the new CTX. Coincidentally, fear conditioning causes CeL_{on} neurons to become CS-responsive. These neurons inhibit CeL_{off} neurons and disinhibit CeM, causing a defensive response. During extinction, CTX^{-} induced safety responses are learned in BLA_{pyr}; these signals compete with the fear signals from LA_{pyr} and CeL_{on} and partially suppress the defensive response in CeM. The decoded output of CeM (last panel) indicates the model's overall defensive response.

transmitted to all three populations via fixed connections, but each population responds to the CS differently. LA_{pyr} is initially unresponsive to the CS, but (some) neurons in this population become CS-responsive during acquisition. Specifically, synaptic weights on the connection between EXT_{cs} and LA_{pyr} are initialized to zero and updated using an the error signal computed in LA_{err}. In contrast, LA_{inh} responds to the presence of CS via a fixed connection from EXT_{cs}: If any CS is present, LA_{inh} will represent $\hat{\mathbf{x}} = 1$ and will otherwise represent $\hat{\mathbf{x}} = 0$.

LA_{err} represents the difference between the US and CS values represented in LA_{pyr}. If this difference (US minus CS) is positive, it means that a US is currently active but that LA_{pyr} is not responding to the coincident CS. To correct this error, we apply PES (Equation 4) to update the decoders on the connection between EXT_{cs} and LA_{pyr}. Online learning is driven by positive error in LA_{err}, causes LA_{pyr}'s CS response to increase and stops when the computed error is reduced to zero. Unfortunately, with this configuration, repeated presentation of the CS without the US will negate the CS-US association: Negative representations in LA_{err} produce negative changes in the decoders that directly oppose the original association. To ensure this unlearning does not occur in later stages of the experiment, we use two additional mechanisms. First, we use LA_{inh} to convey US information to LA_{err}; if we externally inhibit LA_{inh}, we can control neural activities and prevent learning. Second, we set the encoders, gains and biases in LA_{err} such that its neurons are not active when the CS exceeds the US; setting the tuning curves in this way causes the population to represent $\hat{\mathbf{x}} = 0$ whenever the error is negative, which also prevents learning.

Figure 2 (second panel) shows the dynamics of fear acquisition in LA. The y axis plots the values decoded from the spiking activities of simulated neural populations, with higher values indicating a stronger response. We begin with a baseline test: The CS is presented in isolation, but LA_{pyr} does not respond. Next, five CS-US pairings are presented: LA_{pyr} exhibits a large US response but no CS response, leading to a large error in LA_{err} (fifth panel) and rapid updates of the connection between EXT_{cs} and LA_{pyr}. By the fourth pairing, the CS response has risen to meet the US response. To simulate extinction, we switch the context from CTX^{+} to CTX^{-}, then repeatedly apply the CS. Due to the tuning of LA_{err} neurons, the negative error between LA_{pyr}'s US and CS response is encoded as $\hat{\mathbf{x}} = 0$, preventing unlearning and leading to persistent fear.

Next, we add several populations for the CeA. Broadly speaking, CeL learns CS-US associations during acquisition and transmits defensive responses from LA to CeM; while CeM is the output nucleus, combining fear and safety signals from all other populations into a final defensive response. LA_{pyr} connects to CeL_{on}, which inhibits CeL_{off}, which in turn inhibits CeM. During acquisition, CeL_{on} learns to mirror the CS responsiveness of LA_{pyr}, as shown in Figure 2 (third panel): The error population CeL_{err} calculates the difference between responses in CeL_{on} and LA_{pyr}, which drives learning on the connection between LA_{pyr} and CeL_{on}. Following acquisition, CeL_{on} inhibits CeL_{off}, which disinhibits CeM and creates a defensive response.

Finally, we add four BLA populations. BLA_{pyr} represents learned responses to CS, US and CTX; it receives CS and US information from LA_{pyr} and CTX information from EXT_{ctx}. The former connection is fixed, and the latter is learned with the help of $BLA_{err\text{-}safe}$ and $BLA_{err\text{-}fear}$, which govern extinction learning and context-dependent fear conditioning, respectively. BLA_{inh} gates contextual learning in a manner similar to LA_{inh}: It receives CS and US information from BLA_{pyr} and transmits these signals to the error populations; and it inhibits the error populations such that (a) contextual fear acquisition occurs faster than de-acquisition and (b) extinction learning only occurs when the CS is present without the US. Finally, BLA_{pyr} connects to CeM, enhancing or suppressing the overall defensive response.

Figure 2 (fourth panel) shows the dynamics of BLA_{pyr} responses during acquisition, extinction and testing. During acquisition, BLA_{pyr}'s CS response is driven by LA_{pyr}, and BLA_{pyr}'s CTX response increases as it learns to associate CTX^{+} with US. At the beginning of extinction, the context switches to CTX^{-}, a novel five-dimensional vector with a unique set of component features. This novel environment activates a different set of EXT_{ctx} neurons; because CTX^{-} shares few feature with CTX^{+}, most neurons that activate in response to CTX^{-} did not undergo synaptic plasticity during acquisition in CTX^{+}. This effectively resets BLA_{pyr}'s CTX response to zero. As extinction proceeds, newly activated EXT_{ctx} neurons become the basis for a new learned association between CTX^{-}, CS and the absence of US. This association drives BLA_{pyr}'s CTX response to negative values; when this CTX^{-}response is added to the CS-induced defensive response in CeM, the result is suppressed fear expression.

4 | RESULTS

Having established the basic functionality of the model in Figure 2, we now run a series of experiments to assess the model's neural and behavioural realism. To do so, we simulate the model and interpret its outputs as predictions in various experimental contexts. We then validate these predictions against empirical data wherever

possible and suggest future experiments to test the remaining predictions.

4.1 | Dynamics of extinction

We began by characterizing the defensive responses exhibited by the model over the course of training. In this experiment, we simulated an acquisition phase (five CS-US pairings in the acquisition context, CTX^+), an extinction phase (10 unpaired CS in either CTX^+ or a novel context CTX^-) and a test phase (five unpaired CS in the extinction context, CTX^+ or CTX^-). We simulated and trained 10 unique instances of our model; to approximate the variance between individual animals, each model instance had randomized tuning curves and learning rates. We characterized the model's defensive response by recording the decoded value from CeM. Figure 3 plots this response as a function of time, with each point representing a single presentation of the CS. Consistent with empirical observations of fear conditioning and extinction, our model's defensive response increased rapidly during acquisition and plateaued at some maximum value, then decreased slowly during extinction (and subsequent testing) and plateaued at some small positive value. Comparing defensive responses when extinction occurs in different contexts, we observed that extinction training is more effective when performed in a novel context (CTX^-) than in the acquisition context (CTX^+). However, with enough extinction training in the acquisition context, the model will still learn to suppress the CS-US association. Note that changing model parameters, such as the learning rates, contextual distinctiveness and the strength of LA versus BLA outputs, determines the (a) amount of time required for successful extinction and (b) the relative difficulty of extinguishing the defensive response in CTX^+ versus CTX^-. Also note that extinction continues during the test phase, because continued presentation of the CS without US will continue to teach contextual safety associations that further suppress the defensive response.

4.2 | Renewal

Next, we performed a fear renewal experiment: Our protocol, adapted from Lonsdorf et al. (2017), involved an acquisition phase (five CS-US pairings in CTX^+), an extinction phase (10 unpaired CS in CTX^-) and a test phase (one unpaired CS in either CTX^+, CTX^- or a novel CTX^*). We generated a dataset for analysis by simulating and training 10 unique instances of our model. Figure 4 (right panel) reports the results: The model's defensive response is zero before training and is greatest when the model is presented with the CS in CTX^+ during testing. In CTX^-, the defensive response is mostly suppressed, but in a novel CTX^*, it is renewed. To validate these results, we compared our model outputs with behavioural data from animal studies and human experiments. Across conditioning paradigms, fear metrics and study animals, the empirical data show a clear trend: Defensive responses are greatest when the CS is presented in CTX^+, are reduced in CTX^- and are renewed to varying degrees in a novel CTX^* (Antoniadis & McDonald, 1999; Bouton & Bolles, 1979; Hermans et al., 2005, 2006; Lovibond et al., 2000; Vansteenwegen et al., 2006). For example, the left panel of Figure 4 plots the fear response of 28 human participants on an aversive conditioning task, operationalized using self-reported evaluative ratings. This confirms our experimental predictions regarding the magnitude of a defensive response following conditioning and extinction.

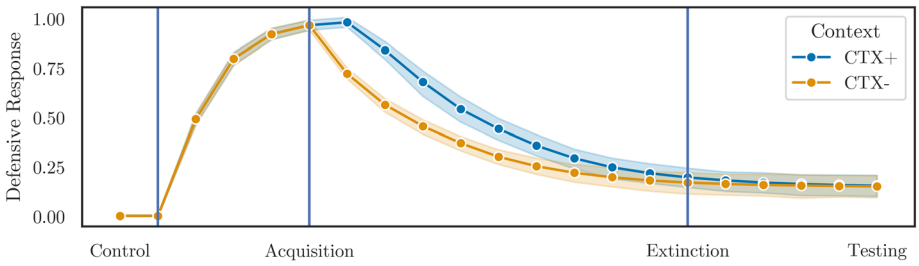

FIGURE 3 Dynamics of defensive responses following extinction in two different contexts. The line represents the defence response averaged across 10 model instances, while the shaded region represents variance between instances. After training a CS-US association during acquisition, we performed extinction training in either the acquisition context (CTX^+) or a novel context (CTX^-). Suppression of the defensive response proceeded faster in a novel context, but given enough training, the model learned to suppress roughly 75% of the acquired defensive response when tested in either extinction context.

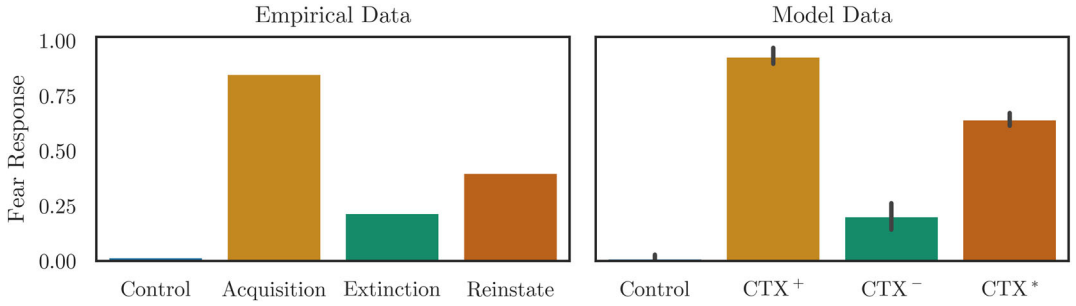

FIGURE 4 Left: Empirical data on fear expression. In Hermans et al. (2005), the authors paired CS^+ with US and CS^- with safety during acquisition, then removed the CS^+-US association during extinction. Reported values indicate the differences between participant CS^+ and CS^- fear responses during a reinstatement test, as measured by self-reporting ($n = 28$ split across two groups; we renormalized response values to fall between 0 and 1). Similar results have been obtained in rats using other protocols (Bouton & Bolles, 1979). Right: Defensive responses across 10 unique model instances. Networks were subjected to acquisition and extinction training then tested in three different contexts: the acquisition context CTX^+, the extinction context CTX^- or a novel context CTX^*. Consistent with empirical data, defensive responses are negligible before conditioning (control) and greatest in CTX^+; defensive responses in CTX^- are greatly reduced due to contextual extinction but reemerge in novel CTX^*. Error bars plot 95% bootstrapped confidence intervals across model instances.

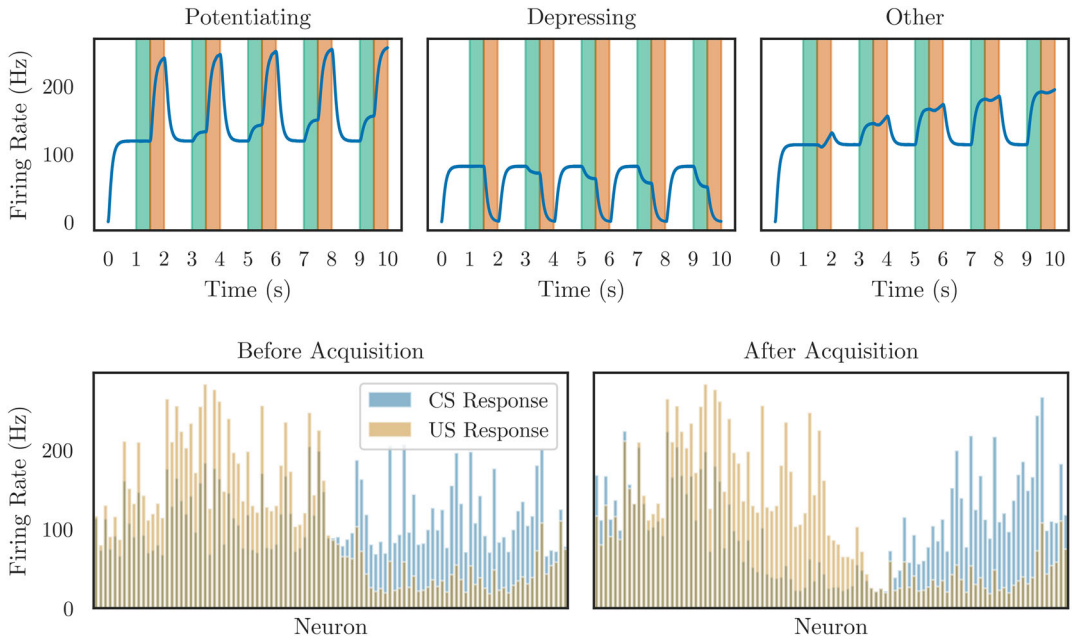

FIGURE 5 Top: Spiking activity of three simulated LA_{pyr} neurons during acquisition. Time windows corresponding to CS and US presentations are shown in green and orange, respectively. Fear conditioning potentiates some neurons (left), depresses others (centre) and may even cause more complex changes (right). Bottom: The CS-induced response of each simulated neuron in LA_{pyr} before (left) and after (right) acquisition training, compared with their US-induced responses before acquisition (orange). In general, the changes in CS representation across the entire population cannot be characterized as movement towards the US representation.

4.3 | CS representation

How does the spiking neural representation of the CS change during the course of learning in our model? To answer this question, we recorded spikes from LA_{pyr} during acquisition and compared them with the responses induced by presenting the US before acquisition. Figure 5 (top panel) shows the firing rate of three neurons in LA_{pyr} over time. The first neuron increases its firing rate in response to US; during acquisition, the CS-induced firing rate of this neuron increases, such that its CS-induced activity begins to resemble its US-induced activity. In contrast, the second neuron lowers its activity in response to US, and during acquisition, its CS-induced activity decreases. Finally, the third neuron is not initially responsive to either US or CS during acquisition, but its

CS- and US-induced activities both increase during training. Fear conditioning thus induces both potentiation and depression among simulated pyramidal neurons, as has been observed empirically in LA and BLA (Grewe et al., 2017).

However, an examination of neural activities across the entire population reveals a more complex story. Figure 5 (bottom panel) plots the CS- and US-induced activities of all simulated neurons in LA_{pyr} before and after acquisition. While a proportion of neurons follow the pattern observed in the top panel (their CS-induced activity changes to resemble their US-induced activity as a result of learning), many neurons do not follow this trend or even exhibit the opposite trend. Our model predicts that learning induces complex changes in the CS response of LA and BLA pyramidal neurons, rather than simply causing these neurons to exhibit US-like responses when presented with the CS, as has been suggested by some (Grewe et al., 2017).

4.4 | Neural responses

To further characterize the response properties of our simulated neurons following acquisition and extinction, we classified neurons in LA and BLA based on their mean firing rates before and after training, as per Ressler and Maren (2019), Bocchio et al. (2017) and Herry et al. (2008). LA_{pyr} neurons were labelled 'up' neurons if their mean activity grew by at least 50% following conditioning (baseline versus CTX^+ test) and labelled 'down' if their mean activity shrank by at least 50%. Similarly, BLA_{pyr} neurons were labelled (a) 'fear' neurons if they became responsive during acquisition but were suppressed by extinction (50% increase from baseline to CTX^+, 50% decrease from CTX^+ to CTX^-); (b) 'extinction' neurons if they became responsive during extinction (50% increase from CTX^+ to CTX^-); and 'persistent' neurons if they became responsive during acquisition and were not suppressed by extinction (50% increase from baseline to CTX^+, then a decrease of less than 50% between CTX^+ and CTX^-).

Figure 6 reports the changes in neural activities following training. Unlike previous AMY models (Carrere & Alexandre, 2015; Mattera et al., 2020; Vlachos et al., 2011), we did not specify the existence of these classes of neurons ahead of time: These response patterns emerge naturally as a result of learning within an undifferentiated and fully connected population of neurons in LA and BLA (we discuss this point further in Section 5.3). The emergence of these characteristic response profiles is qualitatively consistent with electrophysiological data (McDonald, 2020; Krabbe et al., 2018). Our model also quantitatively predicts the relative frequency of neurons from these different classes. For each model instance, we recorded the number of neurons in each category, then divided by the total number of neurons in that population. We found that the 20% to 34% of simulated LA neurons were 'up' neurons while 33% to 39% were 'down' neurons and that 4% to 8%, 25% to 35% and 5% to 15% of simulated neurons were 'fear', 'persistent' and 'extinction' neurons, respectively. To validate these predictions, we searched the literature for empirical estimates of the measured frequency of neurons from each class. Unfortunately, few studies report these metrics, and each study uses a different experimental procedure and measurement criteria (Gründemann & Lüthi, 2015). Still, empirical estimates support our model predictions: For instance, estimates of the prevalence of LA 'up' neurons (those that become more responsive to CS following acquisition) range from 5% at the low end (Kyriazi et al., 2018) to 40% at the high end (Ressler & Maren, 2019),

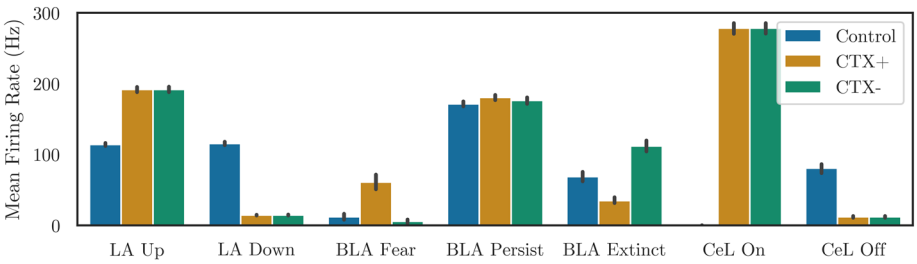

FIGURE 6 Mean firing rates of neurons with characteristic response profiles. Population firing rates were measured by convolving simulated spike trains with a low-pass filter ($\tau = 30$ ms) and averaging over time and neurons. Colouration indicates the context in which these measurements were taken: Blue bars show the baseline mean firing rate before conditioning, while orange and green bars show the mean firing rate following acquisition and extinction training, when the network is presented with the CS in either the acquisition context (orange) or the extinction context (green). For each population, differences between coloured bars demonstrate characteristic response profiles: For example, BLA 'fear' neurons are initially quiet, activate in response to the CS presented in CTX^+ but are inhibited when the CS is presented in CTX^-.

with estimates in the range of 20% to 30% being common (Bocchio et al., 2017; Ressler & Maren, 2019). Future work that rigorously characterizes the frequency of neural responses in LA and BLA, across a large sample of cells and individuals, is needed to further test these predictions.

4.5 | Ablation studies

Many researchers who study fear conditioning use pharmacology, lesioning or direct brain stimulation to externally manipulate neural activities, providing important insights into AMY's functional neuroanatomy. We ran a series of experiments in which we directly inhibited model neurons and observed the effects on acquisition, extinction or expression, thereby predicting the consequence of externally perturbing AMY. Note that these perturbations are not intended to mimic a natural neurobiological process but to recreate empirical 'ablation studies' that help diagnose the functional contribution of each AMY nucleus. In each of the following experiments, we delivered a constant negative current to all neurons within one nucleus during one phase of our experimental protocol, then compared the resulting defence response with the baseline (control) response. Table 1 reports our model predictions and indicates whether empirical data support or contradict each prediction.

Inhibiting pyramidal neurons in LA prevents the learning of CS-US associations during acquisition and reduces fear expression during testing, either in the acquisition context (CTX^+) or the extinction context (CTX^-) (Muller et al., 1997). Our model predicts both these effects. We also predict that inhibiting LA_{pyr} during extinction will impair contextual learning: Although such learning occurs entirely in BLA, our model implies that healthy LA activity is required to relay stimulus information to BLA. We were unable to find any empirical data to validate this prediction.

Inhibiting interneurons in LA prevents the learning of CS-US associations. Empirically, the precise timing of inhibitory spikes is essential for long-term potentiation in pyramidal neurons, and complex networks of inhibition and disinhibition may facilitate pattern separation in LA and BLA (Krabbe et al., 2018). Although PES is not an spike-timing-dependent plasticity rule and does not explicitly perform pattern separation (but see Bekolay, 2011; Bekolay et al., 2013), our model successfully predicts that inhibiting LA_{inh} impairs fear acquisition but preserves fear expression. We also predict that inhibiting LA_{inh} during extinction leads to reduced fear expression, but were unable to find any empirical data for validation.

Inhibiting pyramidal neurons in BLA prevents extinction and leads to stronger defensive responses in CTX^- (Laurent & Westbrook, 2008; Sierra-Mercado et al., 2011). Our model correctly predicts this effect, both for extinction training and subsequent tests in the extinction context. We also observe that inhibiting BLA_{pyr} during acquisition impairs defensive responses, presumably because contextual fear associations are not properly learned. We predict that, in experiments where contextual cues associate with the US, inhibiting pyramidal neurons in BLA will impair acquisition to a modest degree. We note that empirical studies have drawn different conclusions about the effects of inhibiting pyramidal 'BLA' neurons in acquisition versus extinction (Laurent & Westbrook, 2008; Muller et al., 1997; Sierra-Mercado et al., 2011). We attribute this confusion, in part, to the ambiguous boundary between LA and BLA: Application of receptor agonists and antagonists to lateral parts of the AMY will probably affect both CS- and CTX-sensitive pyramidal neurons, which we functionally separate into LA and BLA (respectively) in our model. More

TABLE 1 Defensive responses when one AMY nucleus is externally inhibited during one phase of a fear conditioning protocol.

Inhibited population / During	Acquisition CTX^+	Extinction CTX^-	Expression CTX^+	Expression CTX^-
LA_{pyr}	.007	.802	.347	.0
LA_{inh}	.601	.022	.929	.176
BLA_{pyr}	.527	.710	.385	.385
BLA_{inh}	.527	.636	.929	.176
CeL	.932	-	1.038	-
CeM	.929	-	.0	-
None	.930	.203	.930	.203

Note: Data are collected from 10 unique networks, and mean defensive response is indicated (higher numbers indicate a stronger response). Rows indicate which AMY nucleus was inhibited, and columns indicate when the inhibition was applied. The final row indicate the baseline defensive response when no inhibition was applied (control). Marks indicate whether the model prediction agrees with empirical evidence (green), disagrees (red) or is ambiguous or untested (teal); see text for details.

targeted interventions are therefore necessary to validate some of our predictions for BLA_{pyr}.

Inhibiting interneurons in BLA has varied effects on fear conditioning. Optogenetically inactivating BLA neurons during acquisition appears to facilitate fear learning (Krabbe et al., 2018). Our model predicts the opposite effect: Inactivating BLA_{inh} impairs fear conditioning. One explanation for this discrepancy is that Krabbe et al. (2018) targeted only a single interneuron cell type (PV neurons); inhibiting other interneuron subtypes (such as SOM and CCK) might impair fear acquisition, as we observed. In contrast, disrupting BLA interneurons during extinction, or ablating axoaxonic inhibitory synapses in BLA (Saha et al., 2017), appears to impair contextual safety learning (Krabbe et al., 2018); our model predictions support this conclusion. Finally, our model predicts that inhibiting BLA interneurons should have no significant effect on expression following training and consolidation, but we were unable to find any validating data.

Inhibiting neurons in CeL produces different effects, depending on when the inhibition is applied. Inhibiting CeL before acquisition removes its default inhibition of CeM, causing unconditioned defensive responses, but inhibiting CeL after acquisition has no significant effect (Ciocchi et al., 2010). Our model correctly predicts both these effects. However, while our model predicts that inactivating CeL during acquisition has no effect, inhibiting CeL during acquisition has been shown to impair fear learning (Ciocchi et al., 2010). This is due to spurious CeL learning during simulated extinction that should be addressed in future versions of the model. Finally, inhibiting CeM impairs fear expression but does not affect acquisition (Ciocchi et al., 2010); our model correctly predicts these results.

4.6 | Fear generalization

To investigate fear generalization, we trained our model by presenting 10 pairs of CS^+-US and 10 unpaired CS^- (order randomized), then presented a series of novel stimuli (CS^*). Recall that we represent CS inputs using high-dimensional vectors that encode complex sensory information; to generate CS^- and CS^*, we simply created more vectors using our original sampling procedure. The similarity between these vectors and CS^+ was calculated using cosine similarity. Figure 7 (right panel) shows how the model's defensive responses changes as a function of this similarity: Our model predicts that defensive responses decrease as CS^* becomes more dissimilar to CS^+ and that this fear 'gradient' appears to follow a sigmoid curve.

We also investigated the hypothesis that fear gradients depend on pattern separation in AMY, sensory cortex and/or hippocampus. According to this theory (Dunsmoor & Paz, 2015; Lissek et al., 2014), the overgeneralized fear responses observed in anxiety disorders occur when neurons in these areas are insufficiently selective for the CS^+ (i.e., they continue firing even when presented with dissimilar stimuli). To test the idea that poor pattern separation causes broader defensive responses to novel CS^*, we modified the tuning properties of our neurons and repeated the above experiment.

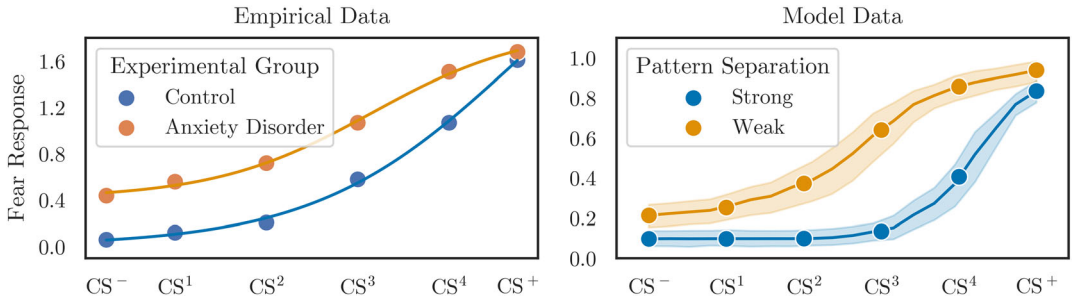

FIGURE 7 Left: Empirical data on fear generalization. In Lissek et al. (2014), the authors paired the presentation of a large circle (CS^+) with a mild electric shock and paired the presentation of a small circle (CS^-) with no shock. They then presented circles of intermediary sizes (CS^1–CS^4) and measured participant defensive responses using a self-reported risk rating (or startle EMG, not shown). They performed this experiment in both healthy controls ($n = 26$) and patients with generalized anxiety disorder (GAD, $n = 22$). They found that generalization gradients were more gradual in GAD patients. Sigmoids were fit to the data using Scipy's *curve_fit* function (Virtanen et al., 2020). Right: Simulated fear across 10 unique model instances. To manipulate the degree of pattern separation in our model, we adjusted the tuning curves in EXT_{cs} to make neurons more (or less) sensitive to a broad range of input stimuli. We found that models with stronger pattern separation exhibited sharper fear generalization gradients, while networks with weaker pattern separation exhibited more gradual gradients. This trend is consistent with empirically observed differences between healthy humans and individuals with anxiety disorders. Shaded regions indicate 95% bootstrapped confidence intervals.

In our model, external stimuli are first represented in EXT_{cs} and EXT_{ctx}, which correspond to sensory cortex and hippocampus. When initializing these populations, we use encoders that evenly tile the input space, ensuring that all possible stimuli engage a subset of neurons. We also specify the sparsity of neural representation in EXT_{cs} by setting α_i and β_i, which effectively determines how similar a CS input must be to the encoder $\mathbf{e}_i$ before the neuron begins firing. By varying the sparsity of neurons in EXT_{cs}, we can thus control the degree of pattern separation in our model. We predicted that sparse networks would exhibit narrow fear gradients, while broadly tuned networks would exhibit extended fear gradients.

Figure 7 shows generalization gradients for two sparsity values. For each sparsity, we trained 10 unique model instances and tested each with a set of 20 CS^*s. We observed that models with sparser neural representations (stronger pattern separation) displayed sharper generalization gradients. In sparse networks, defensive responses declined rapidly as soon as CS^* became distinguishable from CS^+. In contrast, in networks with weak pattern separation, defensive responses remained high for CS^*s that were similar to CS^+ and only declined as CS^* began to resemble CS^-.

Several empirical experiments have characterized how fear responses to CS^* decreases with similarity to CS^+. Figure 7 (left panel) reproduces the data from an influential experiment, which studied fear generalization in both healthy control subjects and participants with generalized anxiety disorder (GAD) (Lissek et al., 2014). The authors found that the fear experienced by control subjects, operationalized using either a self-reported risk rating or startle EMG, declined sharply as similarity to CS^+ decreased, but that the gradient for GAD subjects was more gradual. In healthy controls, fear declined rapidly between CS^+ and CS^2 then plateaued for CS^1 and CS^-. In contrast, in GAD participants, fear remained high between CS^+ and CS^4, then declined gradually until CS^-. Similar gradients are apparent in patients with panic disorders (Lissek et al., 2010) and PTSD (Lissek & van Meurs, 2015). These results are consistent with our predictions, and with the hypothesis that GAD patients overgeneralize learned fear responses, reacting strongly when CS^* shares even a few features with the CS^+ (Duits et al., 2015).

Another feature of the empirical data is that anxious individuals appear to have a heightened fear response to CS^-.[3] However, our model predicts that the defensive response to CS^- should be similar in networks with strong versus weak pattern separation (not shown). To account for this difference, we hypothesized that anxious individuals also learn stronger associations between CTX and US, leading to higher baseline defensive responses when testing in the acquisition context. We tested this idea by strengthening the connections between BLA and CeM in our model, causing contextual fear learning to have a greater affect on the overall defensive response. When we repeated the generalization experiment, we observed stronger defensive responses for all CS, as shown in Figure 7 (right panel, orange line). This result supports the idea that the shape of the generalization gradient (slope and inflection point) is determined by pattern separation in AMY and hippocampus but that baseline levels of fear following acquisition (y intercept) depends on intrinsic connectivity within AMY.

To confirm our characterization of fear generalization and to demonstrate the scalability of our model to more complex stimuli, we repeated this generalization experiment for multiple values of the CS dimensionality (D) and network sparsity. In order to visualize the entire generalization gradient, we trained the model with a CS^- that was maximally dissimilar to CS^+, then tested 100 CS^*s with intermediate similarities. Figure 8 shows that generalization gradients appear sigmoidal regardless of stimulus complexity and that weakening pattern separation in the model consistently shifts this sigmoidal gradient to the left. Unfortunately, the empirical data on fear generalization are too sparse to robustly validate our findings: The current data all use the same experimental paradigm, do not report error bars in the plots and use only simple stimuli.

Drawing on the insights gained from the simulated experiments in Figures 7 and 8, we conclude with three predictions for future experiments on fear generalization. First, we predict that the relationship between defensive response and stimulus similarity will be well characterized by a sigmoid curve. Second, we predict that the generalization curves of anxious individuals will be shifted left (with centres towards more dissimilar stimuli) relative to healthy individuals. And third, we predict that experiments that use simpler stimuli (those that include few discernible features) will produce more extreme differences between healthy and anxious individuals (larger shifts to the left).

5 | DISCUSSION

In this paper, we presented a spiking neuron model of the AMY, subjected it to several fear conditioning protocols and compared its predictions to neural and

[3]Note that the empirical data from Lissek et al. (2014) does not include a measure of variance, so we should be cautious when drawing quantitative conclusions about generalization gradients in anxious versus healthy humans.

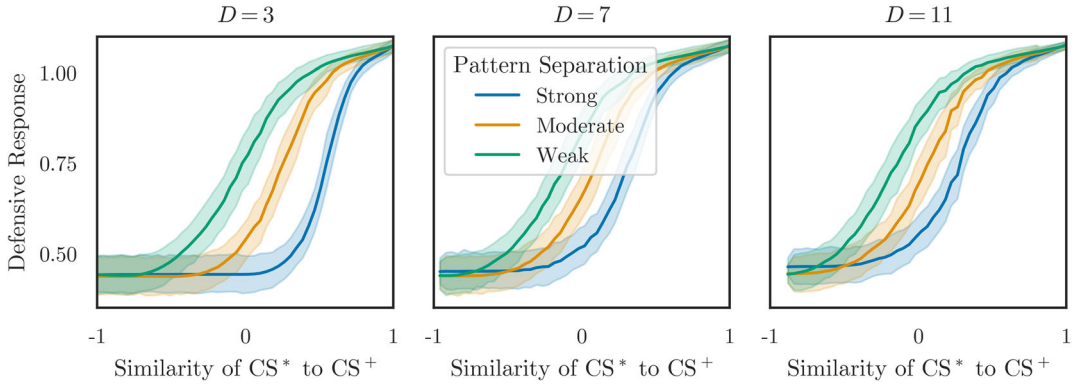

FIGURE 8 Simulated fear generalization as a function of stimulus complexity and network sparsity. As the dimensionality D of the input stimulus increases, fear gradients maintain their sigmoid shape, regardless of the degree of pattern separation in the model. The gradients of networks with weaker pattern separation are also consistently shifted to the left. However, differences between the observed gradients in networks with strong versus weak pattern separation are most obvious for low-complexity stimuli ($D = 3$) and are less apparent for high-complexity stimuli ($D = 11$).

behavioural data from the empirical literature. Our hypothesis was that fear conditioning and extinction result from error-driven learning of the synaptic weights between specific AMY nuclei. In this section, we discuss the various assumptions made by our AMY model, review the predictions we made with regards to existing data and future experiments, compare our model to other computational models of learning in AMY and identify meaningful ways to extend the model in future work.

5.1 | Functional capacity

Our model boasts three key functional features. First, our model clearly specifies the functional roles of numerous populations and connections within AMY; that is, it explains how fear learning emerges from specific neural representations and computations within AMY. This functional capacity was realized using the NEF: Drawing on previous theoretical accounts of learning within AMY (Carrere & Alexandre, 2015; Duvarci & Pare, 2014; Mirolli et al., 2010), we created a model in which each nucleus plays a particular role within a larger AMY learning system (Section 2.2 and 2.3). Our model thus helps translate the high-level ideas put forward by previous theorists into a quantitative model that makes falsifiable predictions.

Second, our model includes only as many biological features and computational mechanisms as is required explain the neural and behavioural results that we examine. This constraint relates to a fundamental trade-off faced by all computational models of the brain: Should they include as many biological features as possible, in order to better resemble the organization of the brain? Or should they seek to explain how a particular cognitive ability arises from a particular set of biological features? Our model takes a middle-of-the-road approach: We do not include extra biological features for complexity's sake (each biological feature in our model has a functional purpose), but we identify a diverse set of cognitive abilities and neural results that we wish to reproduce. We found that in order to explain all these cognitive abilities and neural patterns within a single model, we had to include a wide set of interacting biological features. We thus consider our model (a) functional, in the sense that it requires all of the simulated components to produce the desired results, and (b) biologically plausible, in the sense that it leverages many known biological features. In Sections 5.2 and 5.3, we discuss in detail the biological features that were included versus excluded in the model and how they extend the functionality of previous neural models.

Third, our model simulates high-dimensional representations and makes novel predictions about fear generalization. In order to simulate complex external inputs and explain how AMY represents and transforms this information, we used a two-fold approach. First, our model represents CS and CTX as higher dimensional vectors that drive spiking activity in our neural populations, effectively encoding multiple features of the external world. Second, we use an online, spike-based, error-driven learning rule (PES) to realize associative learning in these high-dimensional spaces. The conjunction of high-dimensional representations and error-driven learning gives our model several unique functional properties. Most importantly, PES leverages knowledge about the representations of model neurons to update synaptic weights based on their contribution to the current error. This preserves information about the features of high-dimensional inputs as they are collapsed into

low-dimensional 'salience' representations in LA and BLA. One consequence is that these salience responses generalize to similar inputs: As CS and CTX change, salience responses change proportionally, leading to a smooth generalization gradient. Furthermore, PES naturally produces several classes of neural responses observed in LA and BLA: Although all pyramidal neurons in LA and BLA receive the same stimulus information, their CS responses naturally diverge as a result of learning.

These three functional features produce simulated behaviour that both reproduces known behavioural results in the fear conditioning literature and makes novel predictions for future experiments. During acquisition, we placed the model in CTX^+ and presented paired CS and US: The model learned that CS and CTX^+ predicted the onset of the US and produced the defensive response to CS or CTX^+ when these stimuli were presented without the US. During extinction, we presented the CS without the US in a new CTX^-: The model learned that CTX^- was associated with safety from the US and suppressed its CS-induced defensive response when tested in CTX^-. However, when the model was returned to CTX^+, or placed in a novel CTX^*, it exhibited fear renewal, confirming that extinction-induced fear suppression is context specific. The model's degree of defensive response in CTX^+, CTX^- and CTX^* is consistent with animal experiments (Hermans et al., 2006; Vansteenwegen et al., 2006): Responses in the extinction context are suppressed, responses in the acquisition context remain high and responses in novel contexts are intermediate. We also showed that the model learned to suppress its defensive response when acquisition and extinction took place in the same context, albeit at a slower rate.

Our model generalized fear associations to similar stimuli and contexts. After showing that the model produced a defensive response to one particular input vector (CS^+), we investigated how this association generalized to similar vectors in the input space (CS^*). We found that when neural representations were sufficiently sparse (i.e., realized strong pattern separation), our simulated gradients agreed with empirical gradients (Dunsmoor et al., 2009; Lissek et al., 2008): Both simulated and empirical defensive responses declined rapidly as similarity between the CS^* and CS^+ decreased. This result is consistent with generalization gradients across animal groups, behavioural contexts, sensory modalities and learning styles (Ghirlanda & Enquist, 2003). Interestingly, we found that when we decreased the sparsity of neural representations (i.e., realized weak pattern separation), simulated defensive responses declined more gradually, remaining high until CS^* began to resemble CS^-. These results resemble the gradients observed in patients with fear-related disorders, including generalized anxiety disorders, post-traumatic stress disorder and panic disorders (Duits et al., 2015; Dunsmoor & Paz, 2015; Lissek et al., 2014, 2010, Lissek & van Meurs 2015). Our computational model thus provides novel theoretical support for the idea that anxiety disorders are mediated by weaker pattern separation in areas such as MTL, hippocampus and AMY and quantitatively predicts the shapes of these curves in healthy versus anxious individuals.

Despite these successes, our model is not without its functional limitations. In this paper, we simulated input signals (CS, US and CTX) that contained multiple features but did not include temporal complexity. Our learning rule requires that CS and US be coincident for association to occur. However, in many fear conditioning protocols, the CS consists of a temporal sequence of sensory cues (such as a train of audio cues at a certain frequency), and the US often follows the offset of the CS. Fortunately, our modelling framework can accommodate these more realistic stimuli. Previous work, both empirical (Hall et al., 2002; Jones et al., 2007; Kumar et al., 2014) and theoretical (Bendor, 2015; Krishnan et al., 2014), suggests that cortex and hippocampus classify temporal sequences into coherent representations that can be manipulated by other neural systems. Because our model already operates over high-dimensional representation, we can simulate neural networks that process temporal sequences of sensory information, bundle this temporal history into a single vector, then send this vector into the current AMY model. Previous NEF models have demonstrated the feasibility of this approach (Dumont et al., 2023). As for capturing delays between CS offset and US onset, we need only add a short-term memory into the model, either outside AMY (working memory buffers in cortex) or within the respective AMY nuclei. Again, previous NEF models have used biologically plausible neural integrators to build working memories for high-dimensional representations (Duggins & Eliasmith, 2022) and shown that they can be used in large-scale neural networks to perform cognitive tasks (Eliasmith, 2013). Overall, our model should naturally scale to inputs with greater temporal complexity.

A more challenging issue for our model is the consolidation of fear memories. A wide range of neural and behavioural data suggest that, in the hours and days following acquisition and/or extinction, processes occur that solidify learned associations (McGaugh, 2000). Accounts about the synaptic processes underlying consolidation differ widely, from gene regulation (Ressler et al., 2002) to GABAergic neurotransmission (Makkar et al., 2010) to oscillatory synchrony (Totty et al., 2017) to interneuron plasticity (Lucas & Clem, 2018). These processes seem to

involve the coordination of multiple brain areas like AMY, hippocampus and cortex (Chaaya et al., 2018; Feng et al., 2014; Marek et al., 2013) over an extended period of time (Igaz et al., 2002; Pace-Schott et al., 2015). The behavioural consequences of consolidation are also complex: Reversal of the CS-US association may be possible before consolidation (Myers et al., 2006; Schiller et al., 2008)); and effects such as reinstatement and spontaneous recovery may depend on the consolidation of extinction memories. Given the neural complexity of consolidation, its reliance on the coordinated interaction of other brain areas with AMY, and its poorly understood behavioural signatures, we have chosen not to investigate consolidation in the current work. As a result, our model does not offer explanations for many consolidation-related phenomena: For instance, it does not exhibit fear reinstatement or recovery. However, future work could extend our model to investigate the neural mechanisms of fear consolidation. Currently, our model uses interneuron populations in LA and BLA to control the rate and timing of fear acquisition and extinction. Externally modulating these neurons, for instance, via consolidated associations in hippocampus, may result in some of the neural and behavioural signatures of consolidation. Alternately, synaptic plasticity on interneurons themselves may drive consolidation: An expanded version of our model could investigate the effects of learning in these populations, in addition to pyramidal neurons. Unfortunately, given the theoretical and biological complexities of such an undertaking (see below), these investigations are beyond the scope of the current work.

5.2 | Biological plausibility

As mentioned above, our model includes only those biological features that were necessary to explain several well-known results in fear conditioning. Nevertheless, we argue that the model is biologically plausible, because it (a) includes numerous biological features that are not typically investigated in computational models focused on fear behaviour, (b) quantitatively explains how these features contribute to fear learning and (c) predicts several biological results that are outside the scope of existing neural models. In this section, we review the biological features we have included in the model, discuss our biological predictions and indicate additional features that should be incorporated into the model in future work.

We focus on fear learning within AMY: While many brain structures contribute to fear conditioning, contextual extinction and memory consolidation (Luchkina & Bolshakov, 2019), AMY is a key structure in this process. Recent meta-analyses have shown that AMY is one of, if not the, most commonly referenced area in fMRI studies of fear conditioning (Sehlmeyer et al., 2009), and review studies have widely implicated the AMY within larger brain networks for fear conditioning, including prefrontal cortex, hippocampus, IL and PL (Milad & Quirk, 2012). Thus, while AMY is not the sole driver of fear learning in the brain, and while a model of AMY alone cannot explain all fear-related phenomenon, developing models of learning within AMY is indispensable for an understanding of fear learning throughout the brain. Furthermore, the insights we gain regarding representations, computations and learning rules within AMY may be applicable to other brain areas implicated in fear learning. Nonetheless, we acknowledge that numerous other areas contribute to animal fear responses and that understanding fear behaviour in complex animals (especially humans) requires a whole-brain perspective.

Our model simulates two brain areas outside AMY and four nuclei within AMY. Because our focus is on learning within AMY, we do not devote great attention to specifying the biological properties or processes of these two external populations: We simply say that these populations represent information related to context and sensory inputs and that they connect to particular nuclei within AMY (see Section 2.2). With regards to the AMY itself, our model simulates the LA, BLA, CeL and CeM. Within each nucleus, we simulated independent populations of excitatory pyramidal neurons and inhibitory interneurons that communicate via spikes and are connected in an anatomically-plausible manner. The existence and connectivity of pyramidal neurons and interneurons within each nucleus is well established, and our model of their functional connectivity (Figure 1) is derived from Duvarci and Pare (2014). Our model also simulates several neural populations that explicitly encode error signals; while several studies have found evidence of error signals encoded in AMY activity (Li & McNally, 2014; McHugh et al., 2014), error neurons in AMY have not yet been clearly identified, and their existence constitutes a prediction made by our model.

The inclusion of these biological features permits our model to make several novel biological predictions. First, our model produces neurons whose activities change in response to the CS and CTX following acquisition and extinction, including fear, extinction and persistent neurons in BLA. These response profiles emerge naturally as a result of learning: Their existence is not guaranteed by the structure of the model. However, despite the fact that external CS (CTX) inputs are sent to all neurons in LA (BLA) via synaptic connections from EXT_{cs} (EXT_{ctx}), only a fraction of our simulated neurons develop these response properties. We used our model to predict the

relative frequency of these neurons within each nucleus and found that our estimates were broadly consistent with measurements made in empirical experiments (Bocchio et al., 2017; Herry et al., 2008; Ressler & Maren, 2019). Unfortunately, quantitative validation of these figures is difficult in practice: Empirical studies often fail to report the criteria they use to classify neurons and rarely take single-unit recordings (of the same neuron across baseline, training and testing) for many neurons within a given population. Better electrophysiological data are needed to rigorously validate our predictions regarding the abundance of fear, extinction and persistent neurons.

The biological substrate of our model also allows 'ablation studies', in which we artificially inhibit a single AMY nucleus during acquisition, extinction or expression, then predict the subsequent effects on neural activities and fear behaviour. We found that, in almost all cases, our simulated perturbations produced behavioural effects that aligned with empirical experiments, in which chemicals like muscimol, or neurotransmitters like oxytocin, were used to inhibit neurons within specific nuclei (Ciocchi et al., 2010; Krabbe et al., 2018; Muller et al., 1997; Saha et al., 2017; Sierra-Mercado et al., 2011). These results support our theoretical account of how biology and anatomy relate to function and hint at the possibility of investigating neuromodulation in future work. For simulated experiments where validation data could not be found, we made predictions about the effects of inhibiting AMY populations that can be tested in future empirical work.

Future work can extend the biological realism of our model in several directions. We simulate spiking leaky-integrate-and-fire (LIF) neurons; while these neurons provide a better correspondence to biology than mean-field approximations or static nonlinear neurons, they do not exhibit transient bursting (Pendyam et al., 2013), habituation or expectation (Moses et al., 2007). We also do not simulate any physiological differences between pyramidal neurons and inhibitory interneurons: Both are instances of the same LIF model class, use synapses with identical time constants and do not enforce Dale's principle. While such biological abstractions are commonplace in functional AMY models (Carrere & Alexandre, 2015; John et al., 2016; Mannella et al., 2008; Vlachos et al., 2011), we plan to apply recently developed NEF tools (Duggins & Eliasmith, 2022; Stöckel, 2022) to simulate more complex biophysical mechanisms in future work.

Another avenue for model extension relates to the role of inhibitory interneurons, which are known to play an important role in fear learning (Lucas & Clem, 2018; Krabbe et al., 2018). Inhibitory neurons may (a) provide a 'brake' that prevents runaway excitation in pyramidal neurons and controls the magnitude of learning, (b) be a site of synaptic plasticity that contributes to fear or extinction memories, (c) facilitate consolidation of extinction memories during experience replay or (d) encourage pattern separation within AMY nuclei. In our model, interneurons inhibit pyramidal neurons and control learning at pyramidal synapses but do not undergo plasticity themselves. While some other computational models include interneuron populations (Li, 2017; Mattera et al., 2020), these models tend to focus on how anatomical connectivity produces empirically observed patterns of neural activity and do not explore how this connectivity drives learning and behaviour. We are unaware of any computational models that mechanistically explain how inhibitory neurons accomplish (b–d); specifying and modelling the functional role of interneurons is therefore an important topic for future research but requires significant theoretical development that is outside the current scope.

Finally, it is important that future work connect AMY to other brain structures that are implicated in fear learning. For instance, reciprocal connections between AMY and MTL, hippocampus, IL and PL appear to be important for contextual representation and memory consolidation (Milad & Quirk, 2012; Quirk & Mueller, 2008), especially with regards to network oscillations in particular frequency bands (Bocchio et al., 2017; Vlachos et al., 2011). While recurrent connections between AMY, IL and PL have been featured in some low-level models (Mattera et al., 2020; Pendyam et al., 2013), we found that such connections were not necessary to produce fear acquisition, extinction or generalization in our model. However, future work that investigates memory consolidation from AMY to hippocampus and cortex should model the functional roles of the connections between these areas. Such work is eminently possible using the tools we presented in this paper: Recent NEF models have used hippocampal representations and network oscillations to perform navigation tasks guided by contextual cues (Dumont et al., 2023, 2022; Stöckel, 2022). It would be interesting to integrate these mechanisms into our model, with the goal of connecting complex contextual representation in MTL with associative learning in AMY. Lastly, various lines of research point to the role of the BNST in controlling fear expression (Bauer, 2015), conditioning to long-duration stimuli that are associated with anxiety (Waddell et al., 2006) and mediating fear generalization (Duvarci et al., 2009). Unfortunately, theoretical and computational accounts of how BNST interacts with AMY to facilitate fear conditioning are underdeveloped (Poulos et al., 2010; Sullivan et al., 2004), so it is unclear how future work should integrate this structure into the model presented here.

5.3 | Comparison with other computational models

Our model extends previous computational models of AMY fear conditioning in several respects. First, our model simulates numerous biological features that are ignored in simple models of fear learning. Unlike models that investigate the interaction of a few idealized neural populations (Carrere & Alexandre, 2015; Mannella et al., 2008; Moustafa et al., 2013; Vlachos et al., 2011), we simulate (a) multiple nuclei whose internal and external connectivity is anatomically constrained, (b) dedicated populations of inhibitory neurons that functionally control the dynamics of fear conditioning and (c) spiking LIF neurons instead of rate neurons whose firing rates are given by mean-field approximations or static nonlinearities. Despite these added biological features, our model successfully predicts many of the same behavioural phenomenon in fear learning as do the simpler models, including the extent of fear expression following acquisition, extinction and renewal. Furthermore, the scalable nature of our model permits an investigation of fear generalization to novel stimuli and allows us to make novel predictions about the gradient of defensive responses in both healthy and anxious individuals. These experiments cannot be conducted using existing computational models, for reasons discussed below.

On the other hand, our model does not include the level of anatomical detail used in some models (Mattera et al., 2020), nor does it simulate the detailed cellular mechanisms that are central to other models (Kim et al., 2013; Li et al., 2009; Pendyam et al., 2013). While these models make important contributions to our understanding of AMY electrophysiology and synaptic potentiation, we emphasize that our model was designed to investigate a different set of questions and to make predictions at a higher level of analysis, than were these models. As such, these models are better suited to the investigation of, for example, how reciprocal connections with IL and PL sustain neural activities within BLA, whereas our model is better suited to the investigation of, for example, how behavioural defensive responses depend on shared features between the stimuli and contexts present during training versus testing. Still, we argue that our model makes important contributions to a low-level understanding of fear learning and helps connect questions about biological features to questions about cognition and behaviour. For example, our model (a) explains why only a fraction of neurons in LA and BLA become sensitive to CS and CTX following acquisition and extinction and (b) demonstrates the behavioural effects of externally inactivating various nuclei during a fear conditioning protocol.

The second way that we extend previous models of fear conditioning is by using sophisticated representations for fear stimuli and environmental contexts. Many AMY models simulate inputs by injecting an external current into model neurons (Kim et al., 2013; Li et al., 2009; Mattera et al., 2020; Pendyam et al., 2013; Vlachos et al., 2011); such inputs can only convey whether a stimulus is present or absent. Other models use 'one-hot' vectors to represent sensory inputs (Carrere & Alexandre, 2015; John et al., 2016; Mannella et al., 2008; Moustafa et al., 2013); unfortunately, these representations do not permit graded comparison between different CS or CTX, because all possible inputs to the network are maximally dissimilar. Our model represents CS and CTX as higher dimensional vectors that drive spiking activity in our neural populations, effectively encoding multiple features of the external world. We also use PES, an online, spike-based, error-driven learning rule, to realize associative learning. While Hebbian rules are commonplace in both biophysical models (Li et al., 2009; Kim et al., 2013; Pendyam et al., 2013) and conceptual models (Mattera et al., 2020; Mannella et al., 2008; Moustafa et al., 2013; Vlachos et al., 2011), they do not directly account for the features that comprise complex stimuli. On the other hand, PES learning facilitates feature-rich associations and, unlike some other error-driven rules used by AMY models (Carrere & Alexandre, 2015; John et al., 2016), can feasibly be implemented in biological networks (Bekolay, 2011, 2013) and can be applied to many cognitive domains (Eliasmith, 2013; Rasmussen, 2014; Voelker, 2015). As mentioned above, PES naturally produces the neural responses observed in LA and BLA. While some computational models boast the emergence of such responses, many of these models organize neurons with different response properties (e.g., neurons sensitive to one CS but not another, sensitive to CS vs. US or destined for fear vs. extinction) into distinct populations that are connected in a carefully engineered manner (e.g., receive different external inputs or are reciprocally connected via an inhibitory microcircuit) (Carrere & Alexandre, 2015; Mattera et al., 2020; Vlachos et al., 2011). The CS-induced responses of these neurons follow directly from this specific connectivity. In contrast, all pyramidal neurons in our model receive the same stimulus information and are fully connected to the appropriate interneurons, but feature-driven PES learning causes their CS responses to diverge during training.

Finally, while most computational models compare their results to only a single class of empirical data, we validate our model predictions using both neural and behavioural data. Models that learn fear and safety association between stimuli often show that learning changes the CS responsiveness of individual neurons (Kim et al.,

2013; Li et al., 2009; Mattera et al., 2020; Moustafa et al., 2013; Pendyam et al., 2013; Vlachos et al., 2011). Some models decode neural activities to show a change in behavioural defensive responses following training (Carrere & Alexandre, 2015; Li et al., 2009), while others investigate the relationship between context and extinction (Carrere & Alexandre, 2015; Moustafa et al., 2013). Yet another class of models (Moustafa et al., 2013; Pendyam et al., 2013; Vlachos et al., 2011) externally inactivate parts of the model and examine the effects on neural activity or behavioural responses. We validate our model against all these data types: We (a) identified neurons whose CS-induced responses changed during our experiments and compared their mean firing rates, (b) decoded neural activities to estimate defensive responses and showed how they differed in various training regimes and test contexts, (c) externally perturbed each nucleus and observed the effects on fear conditioning and (d) presented the model with novel stimuli and studied gradients of fear generalization in healthy and anxious individuals. In all these cases, we used the model to make falsifiable predictions, then we provided citations for empirical studies that support (or contradict) our results. Notably, ours is perhaps the first neural model to characterize fear generalization as an emergent property of fear conditioning over complex stimuli and makes quantitative predictions about fear gradients that can be tested in future empirical experiments.

6 | CONCLUSION

In this paper, we presented a computational model of fear conditioning, extinction, renewal and generalization in the AMY. We hypothesized that each nucleus in AMY represents specific fear-related quantities and that, when trained using an error signal, the connections between them compute the salience of external stimuli and contexts. Our model recreates AMY neuroanatomy, notably the divisions between its nuclei and the response properties of its pyramidal and inhibitory neurons. We trained the model using online, spike-based, error-driven learning rules that update synaptic connection weights in LA, CeL and BLA, generating or suppressing defensive responses as appropriate. We ran numerous experiments on the model, made a series of neural and behavioural predictions and validated our results with empirical data. Specifically, we identified neurons with different emergent response profiles, measured fear expression in various training protocols and test contexts, performed ablation studies, characterized how fear generalizes to novel stimuli and showed that neural pattern separation can explain differences in defensive responses between healthy and anxious individuals. We concluded by summarizing the biological and functional properties of the model and by comparing it with other computational models of AMY.

Future work should expand the functional capacities, biological realism and cognitive integration of our model. Functionally, we would like to explain additional fear-related behaviours including secondary conditioning, spontaneous recovery, reinstatement and rapid reacquisition. Biologically, we could replace our LIF neurons with biophysically detailed neuron models (Duggins & Eliasmith, 2022), then more realistically simulate the injection of pharmacological agents or study neuromodulation of AMY by dopamine, norephinerphine or oxytocin. We could also recreate inhibitory microcircuits with greater fidelity or explore alternative learning rules (Borst et al., 2018), in order to study the role of interneurons in pattern separation and memory consolidation. Cognitively, we hope to embed our model in larger cognitive networks, study the process of fear memory consolidation in hippocampus and cortex and investigate more complex cognitive tasks. In particular, expanding our model to study the relationship between fear association, salience detection, attention and emotional modulation could help clarify how AMY contributes to affective cognition throughout the brain.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: Peter Duggins conceptualized the experiments, wrote software, developed the methodology, curated the data, performed the investigation and formal analysis, created the figures and wrote the original draft. Chris Eliasmith supervised the project, was responsible for project administration and funding acquisition and helped review and edit the manuscript; he also contributed to conceptualization, investigation, formal analysis and visualization throughout.

ACKNOWLEDGEMENTS

This work was supported by the Canadian Foundation for Innovation (52479-10006, C.E.), the Ontario Innovation Trust (35768, C.E.), the Natural Sciences and Engineering Research Council of Canada (261453, C.E.) and the Air Force Office of Scientific Research (FA9550-17-1-0026, C.E.). The funders had no role in study design, data analysis, decision to publish or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have no competing interests to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16338.

DATA AVAILABILITY STATEMENT

The simulation code for the model and experiments, as well as the empirical and model data, can be found in the Github repository.

ORCID

Peter Duggins https://orcid.org/0000-0002-8948-0586

REFERENCES

Antoniadis, E. A., & McDonald, R. J. (1999). Discriminative fear conditioning to context expressed by multiple measures of fear in the rat. *Behavioural Brain Research*, *101*(1), 1–13.

Bauer, E. P. (2015). Serotonin in fear conditioning processes. *Behavioural Brain Research*, *277*, 68–77.

Bekolay, T. (2011). Learning in large-scale spiking neural networks. (Master's thesis), UWSpace.

Bekolay, T., Kolbeck, C., & Eliasmith, C. (2013). Simultaneous unsupervised and supervised learning of cognitive functions in biologically plausible spiking neural networks. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, 35, pp. 1.

Bendor, D. (2015). The role of inhibition in a computational model of an auditory cortical neuron during the encoding of temporal information. *PLoS Computational Biology*, *11*(4), e1004197.

Bocchio, M., Nabavi, S., & Capogna, M. (2017). Synaptic plasticity, engrams, and network oscillations in amygdala circuits for storage and retrieval of emotional memories. *Neuron*, *94*(4), 731–743.

Boerlin, M., Machens, C. K., & Denève, S. (2013). Predictive coding of dynamical variables in balanced spiking networks. *PLoS Computational Biology*, *9*(11), e1003258.

Borst, J. P., Aubin, S., & Stewart, T. C. (2018). Effective computing in the brain: A whole-task spiking neural network model of associative recognition. In *2018 Conference on Cognitive Computing*, IEEE, pp. 425–428.

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

Carrere, M., & Alexandre, F. (2015). A Pavlovian model of the amygdala and its influence within the medial temporal lobe. *Frontiers in Systems Neuroscience*, *9*, 41.

Chaaya, N., Battle, A. R., & Johnson, L. R. (2018). An update on contextual fear memory mechanisms: Transition between amygdala and hippocampus. *Neuroscience & Biobehavioral Reviews*, *92*, 43–54.

Ciocchi, S., Herry, C., Grenier, F., Wolff, S. B. E., Letzkus, J. J., Vlachos, I., Ehrlich, I., Sprengel, R., Deisseroth, K., Stadler, M. B., & Müller, C. (2010). Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature*, *468*(7321), 277–282.

Conklin, J., & Eliasmith, C. (2005). A controlled attractor network model of path integration in the rat. *Journal of Computational Neuroscience*, *18*, 183–203.

DeWolf, T. (2015). A neural model of the motor control system. (Master's thesis), UWSpace.

Duggins, P., & Eliasmith, C. (2022). Constructing functional models from biophysically-detailed neurons. *PLoS Computational Biology*, *18*(9), e1010461.

Duggins, P., Stewart, T. C., & Eliasmith, C. (2022). Reinforcement learning, social value orientation, and decision making: Computational models and empirical validation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, 44, pp. 1.

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., Van Den Hout, M. A., & Baas, J. M. P. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and Anxiety*, *32*(4), 239–253.

Dumont, N. S. Y., Orchard, J., & Eliasmith, C. (2022). A model of path integration that connects neural and symbolic representation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, 44, pp. 1.

Dumont, N. S.-Y., Stöckel, A., Furlong, P. M., Bartlett, M., Eliasmith, C., & Stewart, T. C. (2023). Biologically-based computation: How neural details and dynamics are suited for implementing a variety of algorithms. *Brain Sciences*, *13*(2), 245.

Dumont, N. S. Y., Furlong, P. M., Orchard, J., & Eliadmith, C. (2023). Exploiting semantic information in a spiking neural SLAM system. *Frontiers in Neuroscience*, *17*, 1190515.

Dunsmoor, J. E., Mitroff, S. R., & LaBar, K. S. (2009). Generalization of conditioned fear along a dimension of increasing fear intensity. *Learning & Memory*, *16*(7), 460–469.

Dunsmoor, J. E., & Paz, R. (2015). Fear generalization and anxiety: Behavioral and neural mechanisms. *Biological Psychiatry*, *78*(5), 336–343.

Duvarci, S., Bauer, E. P., & Paré, D. (2009). The bed nucleus of the stria terminalis mediates inter-individual variations in anxiety and fear. *Journal of Neuroscience*, *29*(33), 10357–10361.

Duvarci, S., & Pare, D. (2014). Amygdala microcircuits controlling learned fear. *Neuron*, *82*(5), 966–980.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*: Oxford University Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*: MIT Press.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, *338*(6111), 1202–1205.

Feng, P., Feng, T., Chen, Z., & Lei, X. (2014). Memory consolidation of fear conditioning: Bi-stable amygdala connectivity with dorsal anterior cingulate and medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *9*(11), 1730–1737.

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron*, *94*(5), 978–984.

Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour*, *66*(1), 15–36.

Gründemann, J., & Lüthi, A. (2015). Ensemble coding in amygdala circuits for associative learning. *Current Opinion in Neurobiology*, *35*, 200–206.

Grewe, B. F., Gründemann, J., Kitch, L. J., Lecoq, J. A., Parker, J. G., Marshall, J. D., Larkin, M. C., Jercog, P. E.,

Grenier, F., Li, J. Z., & Lüthi, A. (2017). Neural ensemble dynamics underlying a long-term associative memory. *Nature*, *543*(7647), 670–675.

Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *12*(2), 140–149.

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

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

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Müller, C., & Lüthi, A. (2008). Switching on and off fear by distinct neuronal circuits. *Nature*, *454*(7204), 600–606.

Igaz, L. M., Vianna, M. R. M., Medina, J. H., & Izquierdo, I. (2002). Two time periods of hippocampal mRNA synthesis are required for memory consolidation of fear-motivated learning. *Journal of Neuroscience*, *22*(15), 6781–6789.

John, Y. J., Zikopoulos, B., Bullock, D., & Barbas, H. (2016). The emotional gatekeeper: A computational model of attentional selection and suppression through the pathway from the amygdala to the inhibitory thalamic reticular nucleus. *PLoS Computational Biology*, *12*(2), e1004722.

Jones, L. M., Fontanini, A., Sadacca, B. F., Miller, P., & Katz, D. B. (2007). Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences*, *104*(47), 18772–18777.

Keifer Jr, O. P., Hurt, R. C., Ressler, K. J., & Marvar, P. J. (2015). The physiology of fear: Reconceptualizing the role of the central amygdala in fear learning. *Physiology*, *30*(5), 389–401.

Kim, D., Paré, D., & Nair, S. S. (2013). Mechanisms contributing to the induction and storage of Pavlovian fear memories in the lateral amygdala. *Learning & Memory*, *20*(8), 421–430.

Krabbe, S., Gründemann, J., & Lüthi, A. (2018). Amygdala inhibitory circuits regulate associative fear conditioning. *Biological Psychiatry*, *83*(10), 800–809.

Krishnan, L., Elhilali, M., & Shamma, S. (2014). Segregating complex sound sources through temporal coherence. *PLoS Computational Biology*, *10*(12), e1003985.

Kumar, S., Bonnici, H. M., Teki, S., Agus, T. R., Pressnitzer, D., Maguire, E. A., & Griffiths, T. D. (2014). Representations of specific acoustic patterns in the auditory cortex and hippocampus. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1791), 20141000.

Kyriazi, P., Headley, D. B., & Pare, D. (2018). Multi-dimensional coding by basolateral amygdala neurons. *Neuron*, *99*(6), 1315–1328.

Laurent, V., & Westbrook, R. F. (2008). Distinct contributions of the basolateral amygdala and the medial prefrontal cortex to learning and relearning extinction of context conditioned fear. *Learning & Memory*, *15*(9), 657–666.

Li, G. (2017). Computational models of the amygdala in acquisition and extinction of conditioned fear, *The amygdalaŮWhere emotions shape perception, learning and memories*. InTechOpen. https://www.intechopen.com/chapters/54746

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

Li, G., Nair, S. S., & Quirk, G. J. (2009). A biologically realistic network model of acquisition and extinction of conditioned fear associations in lateral amygdala neurons. *Journal of Neurophysiology*, *101*(3), 1629–1646.

Lissek, S., Biggs, A. L., Rabin, S. J., Cornwell, B. R., Alvarez, R. P., Pine, D. S., & Grillon, C. (2008). Generalization of conditioned fear-potentiated startle in humans: Experimental validation and clinical relevance. *Behaviour Research and Therapy*, *46*(5), 678–687.

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biological Psychiatry*, *75*(11), 909–915.

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *American Journal of Psychiatry*, *167*(1), 47–55.

Lissek, S., & van Meurs, B. (2015). Learning models of PTSD: Theoretical accounts and psychobiological evidence. *International Journal of Psychophysiology*, *98*(3), 594–605.

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., & Drexler, S. M. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, *77*, 247–285.

Lovibond, P. F., Davis, N. R., & O'Flaherty, A. S. (2000). Protection from extinction in human fear conditioning. *Behaviour Research and Therapy*, *38*(10), 967–983.

Lucas, E. K., & Clem, R. L. (2018). Gabaergic interneurons: The orchestra or the conductor in fear learning and memory? *Brain Research Bulletin*, *141*, 13–19.

Luchkina, N. V., & Bolshakov, V. Y. (2019). Mechanisms of fear learning and extinction: Synaptic plasticity–fear memory connection. *Psychopharmacology*, *236*, 163–182.

MacNeil, D., & Eliasmith, C. (2011). Fine-tuning and the stability of recurrent neural networks. *PloS One*, *6*(9), e22885.

Makkar, S. R., Zhang, S. Q., & Cranney, J. (2010). Behavioral and neural analysis of GABA in the acquisition, consolidation, reconsolidation, and extinction of fear memory. *Neuropsychopharmacology*, *35*(8), 1625–1652.

Mannella, F., Zappacosta, S., Mirolli, M., & Baldassarre, G. (2008). A computational model of the amygdala nuclei's role in second order conditioning. In *International Conference on Simulation of Adaptive Behavior*, Springer, pp. 321–330.

Marek, R., Strobel, C., Bredy, T. W., & Sah, P. (2013). The amygdala and medial prefrontal cortex: Partners in the fear circuit. *The Journal of Physiology*, *591*(10), 2381–2391.

Mattera, A., Pagani, M., & Baldassarre, G. (2020). A computational model integrating multiple phenomena on cued fear conditioning, extinction, and reinstatement. *Frontiers in Systems Neuroscience*, *14*, 569108.

Maurer, A. P., Lester, A. W., Burke, S. N., Ferng, J. J., & Barnes, C. A. (2014). Back to the future: Preserved hippocampal network activity during reverse ambulation. *Journal of Neuroscience*, *34*(45), 15022–15031.

McDonald, A. J. (2020). Functional neuroanatomy of the basolateral amygdala: Neurons, neurotransmitters, and circuits, *Handbook of behavioral neuroscience*, Vol. 26: Elsevier, pp. 1–38.

McGaugh, J. L. (2000). Memory—A century of consolidation. *Science*, *287*(5451), 248–251.

McHugh, S. B., Barkus, C., Huber, A., Capitao, L., Lima, J., Lowry, J. P., & Bannerman, D. M. (2014). Aversive prediction error signals in the amygdala. *Journal of Neuroscience*, *34*(27), 9024–9033.

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

Mirolli, M., Mannella, F., & Baldassarre, G. (2010). The roles of the amygdala in the affective regulation of body, brain, and behaviour. *Connection Science*, *22*(3), 215–245.

Moses, S. N., Houck, J. M., Martin, T., Hanlon, F. M., Ryan, J. D., Thoma, R. J., Weisend, M. P., Jackson, E. M., Pekkonen, E., & Tesche, C. D. (2007). Dynamic neural activity recorded from human amygdala during fear conditioning using magnetoencephalography. *Brain Research Bulletin*, *71*(5), 452–460.

Moustafa, A. A., Gilbertson, M. W., Orr, S. P., Herzallah, M. M., Servatius, R. J., & Myers, C. E. (2013). A model of amygdala–hippocampal–prefrontal interaction in fear conditioning and extinction in animals. *Brain and Cognition*, *81*(1), 29–43.

Muller, J., Corodimas, K. P., Fridel, Z., & LeDoux, J. E. (1997). Functional inactivation of the lateral and basal nuclei of the amygdala by muscimol infusion prevents fear conditioning to an explicit conditioned stimulus and to contextual stimuli. *Behavioral Neuroscience*, *111*(4), 683.

Myers, K. M., & Davis, M. (2002). Behavioral and neural analysis of extinction. *Neuron*, *36*(4), 567–584.

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

Pace-Schott, E. F., Germain, A., & Milad, M. R. (2015). Effects of sleep on memory for conditioned fear and fear extinction. *Psychological Bulletin*, *141*(4), 835.

Pendyam, S., Bravo-Rivera, C., Burgos-Robles, A., Sotres-Bayon, F., Quirk, G. J., & Nair, S. S. (2013). Fear signaling in the prelimbic-amygdala circuit: A computational modeling and recording study. *Journal of Neurophysiology*, *110*(4), 844–861.

Poulos, A. M., Ponnusamy, R., Dong, H.-W., & Fanselow, M. S. (2010). Compensation in the neural circuitry of fear conditioning awakens learning circuits in the bed nuclei of the stria terminalis. *Proceedings of the National Academy of Sciences*, *107*(33), 14881–14886.

Prévost, C., McNamee, D., Jessup, R. K., Bossaerts, P., & O'Doherty, J. P. (2013). Evidence for model-based computations in the human amygdala during Pavlovian conditioning. *PLoS Computational Biology*, *9*(2), e1002918.

Quirk, G. J., & Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology*, *33*(1), 56–72.

Rasmussen, D. (2014). Hierarchical reinforcement learning in a biologically plausible neural architecture. (Ph.D. thesis), UWSpace.

Recanatesi, S., Bradde, S., Balasubramanian, V., Steinmetz, N. A., & Shea-Brown, E. (2022). A scale-dependent measure of system dimensionality. *Patterns*, *3*(8), 100555.

Ressler, R. L., & Maren, S. (2019). Synaptic encoding of fear memories in the amygdala. *Current Opinion in Neurobiology*, *54*, 54–59.

Ressler, K. J., Paschall, G., Zhou, X.-L., & Davis, M. (2002). Regulation of synaptic plasticity genes during consolidation of fear conditioning. *Journal of Neuroscience*, *22*(18), 7892–7902.

Saha, R., Knapp, S., Chakraborty, D., Horovitz, O., Albrecht, A., Kriebel, M., Kaphzan, H., Ehrlich, I., Volkmer, H., & Richter-Levin, G. (2017). Gabaergic synapses at the axon initial segment of basolateral amygdala projection neurons modulate fear extinction. *Neuropsychopharmacology*, *42*(2), 473–484.

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

Sehlmeyer, C., Schöning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., & Konrad, C. (2009). Human fear conditioning and extinction in neuroimaging: A systematic review. *PloS One*, *4*(6), e5865.

Sierra-Mercado, D., Padilla-Coreano, N., & Quirk, G. J. (2011). Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology*, *36*(2), 529–538.

Sotres-Bayon, F., & Quirk, G. J. (2010). Prefrontal control of fear: More than just extinction. *Current Opinion in Neurobiology*, *20*(2), 231–235.

Stöckel, A. (2022). Harnessing neural dynamics as a computational resource. (Ph.D. thesis), University of Waterloo.

Stöckel, A., Stewart, T. C., & Eliasmith, C. (2021). Connecting biological detail with neural computation: Application to the cerebellar granule–golgi microcircuit. *Topics in Cognitive Science*, *13*(3), 515–533.

Stewart, T. C., Tang, Y., & Eliasmith, C. (2011). A biologically realistic cleanup memory: Autoassociation in spiking neurons. *Cognitive Systems Research*, *12*(2), 84–92.

Sullivan, G. M., Apergis, J., Bush, D. E. A., Johnson, L. R., Hou, M. J. E. L., & Ledoux, J. E. (2004). Lesions in the bed nucleus of the stria terminalis disrupt corticosterone and freezing responses elicited by a contextual but not by a specific cue-conditioned fear stimulus. *Neuroscience*, *128*(1), 7–14.

Totty, M. S., Chesney, L. A., Geist, P. A., & Datta, S. (2017). Sleep-dependent oscillatory synchronization: A role in fear memory consolidation. *Frontiers in Neural Circuits*, *11*, 49.

Vansteenwegen, D., Vervliet, B., Hermans, D., Beckers, T., Baeyens, F., & Eelen, P. (2006). Stronger renewal in human fear conditioning when tested with an acquisition retrieval cue than with an extinction retrieval cue. *Behaviour Research and Therapy*, *44*(12), 1717–1725.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ..., & SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272.

Vlachos, I., Herry, C., Lüthi, A., Aertsen, A., & Kumar, A. (2011). Context-dependent encoding of fear and extinction memories

in a large-scale network model of the basal amygdala. *PLoS Computational Biology*, *7*(3), e1001104.

Voelker, A. R. (2015). *A solution to the dynamics of the prescribed error sensitivity learning rule*: Centre for Theoretical Neuroscience.

Voelker, A. R., Crawford, E., & Eliasmith, C. (2014). Learning large-scale heteroassociative memories in spiking neurons, *Unconventional computation and natural computation*, Vol. 7: Springer International Publishing.

Waddell, J., Morris, R. W., & Bouton, M. E. (2006). Effects of bed nucleus of the stria terminalis lesions on conditioned anxiety: Aversive conditioning with long-duration conditional stimuli and reinstatement of extinguished fear. *Behavioral Neuroscience*, *120*(2), 324.

Webler, R. D., Berg, H., Fhong, K., Tuominen, L., Holt, D. J., Morey, R. A., Lange, I., Burton, P. C., Fullana, M. A., Radua, J., & Lissek, S. (2021). The neurobiology of human fear generalization: Meta-analysis and working neural model. *Neuroscience & Biobehavioral Reviews*, *128*, 421–436.

Zilli, E. A. (2012). Models of grid cell spatial firing. *Frontiers in Neural Circuits*, *6*, 16.

How to cite this article: Duggins, P., & Eliasmith, C. (2024). A scalable spiking amygdala model that explains fear conditioning, extinction, renewal and generalization. *European Journal of Neuroscience*, 1–24. https://doi.org/10.1111/ejn.16338