Fear Conditioning and Extinction:
A Computational Model on Interactions Between the Amygdala, Hippocampus and Prefrontal Cortex

(https://github.com/zytyz)

Abstract

Studies have shown that "cognition" and "emotion" partner up in decision-making, and the absence of either of them in the human brain may lead to inappropriate of irrational responses. However, how cognition and emotion are processed in the human brain and the interactions between underlying circuits in the human brain still remains unclear. In order to shed light on the problem, the roles of the amygdala and prefrontal cortex must be analyzed. In this report, I proposed a computational model for the interactions between the prefrontal cortex, the amygdala, and the hippocampus. The model is able to simulate the "fear conditioned" response are results are also evaluated.

1. Introduction

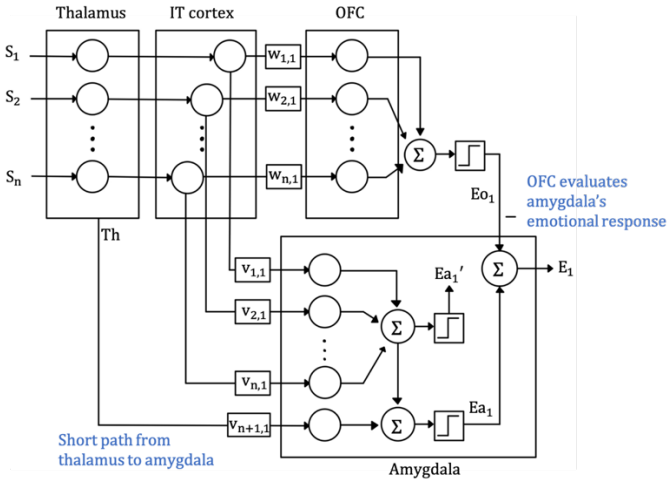
1.1 Cognition and Emotion: Cooperation instead of Antagonism

While cognition is often equated as rationality and logic, emotion, on the other hand, is often mistaken as the cause of irrational behaviors. This idea that cognition and emotion are separable and competitors for dominance of the human brain may sound reasonable and intuitive at first. However, lots of evidence has shown that emotion and cognition are both required for a human being to make the correct decision. For instance, a pathological absence of emotions leads to profound impairment of decision making [2]. In fact, neuroanatomical studies reveal that brain areas responsible for cognitive or emotional processes, which are the prefrontal cortex and the amygdala respectively, are linked in ways that imply complex interactions [1]. Hence, this report aimed to understand this interaction between the cognitive and emotional brain areas.

1.2 The previously proposed model lacks local circuits of the amygdala

The BELPIC model is a model about emotion learning which simulates the roles of the amygdala and the prefrontal cortex during emotional response [4]. The structure of the model is shown in (Fig. 1), and the details for the model can be found in my previous model. Briefly speaking, an emotional stimulus is first passed to the amygdala through a short path from the thalamus. The stimulus is also passed through another pathway that initiates from the thalamus, passes by the IT cortex, and terminates at the orbitofrontal cortex, which evaluates if the emotional response of the amygdala is appropriate. Although the structure of the model does consider the

interaction between the emotion (amygdala) and cognition (orbitofrontal cortex), the amygdala is still treated as a "black box", without further details for the underlying structure and biology basis of the amygdala. Hence, we wish to propose a new model that discusses different regions of the amygdala and their corresponding roles. Although BELPIC simulates all six basic emotions, for the following of this report, I particularly chose the basic emotion "fear", specifically "conditioned fear", as my aim since fear is the most commonly researched emotion of all. The theory and definitions of conditioned fear is introduced in the following section.



(Fig. 1) The structure of the BELPIC model.

2. Theories for Conditioned Fear

2.1 Classical Conditioning

Classical conditioning (also known as Pavlovian conditioning) is learning through association and was discovered by Pavlov, a Russian physiologist. In the process of learning, two stimuli are paired up together while only one of them actually triggers a response, which we called as an unconditioned response (UR). This stimuli is called the unconditioned stimuli (US), while the other neutral one is called the conditioned stimuli (CS). The response could be in any type, emotional responses or behaviors etc. However, as the two stimuli are constantly paired up, the human brain would eventually associate the two stimuli and produced a learned response which we call as the conditioned response (CR). That is to say, when only the conditioned stimuli is shown, it would actually be able to trigger the same response. This learning process of pairing up the US and CS is called classical conditioning. If the learned response of classical conditioning is a fear response, the process is also known as "conditioned fear".

2.2 Conditioned Fear: Acquisition, Extinction, Reacquisition, and Renewal

There are different types of conditioned fear, depending on how the response is learned, which are namely fear acquisition, fear extinction, fear reacquisition, and fear renewal. Fear acquisition refers to the increased expression of the CR as a result of CS-US pairing. Fear extinction refers to the reduction of CR expression when the CS is no longer paired with the US. For example, consider a mouse in a box. A tone appears every time when the mouse experiences a foot shock and thus freezes. Eventually the mouse freezes upon hearing the tone even without a foot shock, implying a fear acquisition process. If foot shocks stop appearing after the tone, eventually the mouse will not freeze upon the tone, which is a fear extinction process.

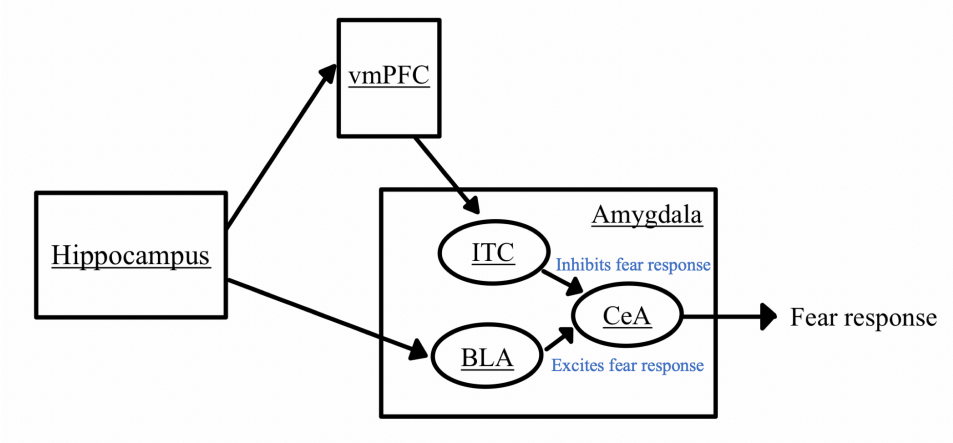
2.3 Contextual and Cue Conditioning

So far, the fear conditioning we have discussed from the mouse model is called "cue conditioning", where the originally neutral stimuli (CS) is the tone. Studies have shown, however, the tone is not the only stimulus that the mouse associates with the foot shock. If the mouse was placed in a box with a certain odor and some other features, the mouse might link the foot shock to the environment of the box as well, showing identical freezing fear responses when placed in the same box again. In this situation, the originally neutral stimuli also includes the features and odors in the box, which we refer to as "context". The conditioned fear triggered by "context" (the features of the box) rather than "cue" (tone) is called "contextual conditioning".

3. Neurobiology Basis Behind Conditioned Fear

3.1 The Amygdala, Prefrontal Cortex, and the Hippocampus

Evidence has shown that three different brain areas have been implicated in fear conditioning: amygdala, hippocampus, and the ventral-medial prefrontal cortex (vmPFC). The connections between these brain regions are depicted in (Fig. 2) [5].



(Fig. 2) Connectivity between the hippocampus, vmPFC, and the amygdala

3.2 The Amygdala Subsystems

The amygdala is a collection of different nuclei, including the central nucleus (CeA), lateral and anterior basolateral nuclei (BLA). CeA is responsible for the initiation of fear response, as the neurons projects to parasympathetic nervous system, hypothalamus, and brainstem motor areas. These areas controls heart rates, freezing, and release of stress hormones, which are all categorized as fear responses. CeA receives projections from both BLA and intercalated cells (ITC), which is also part of the amygdala. BLA neurons are excitatory and facilitates fear responses from the CeA while ITC inhibits fear responses. Hence, it can be said that the BLA is responsible for fear acquisition while the ITC is responsible for fear extinction.

3.3 The Role of Hippocampus

As it was mentioned before in 2.3, conditioned fear can be categorized into "contextual" and "cue" conditioning. Studies have shown that only contextual conditioning is hippocampus dependent, and both cue and contextual conditioning is amygdala dependent [6]. This implies not all conditioned fear is related to the hippocampus and that the hippocampus plays a role in contextual conditioning by storing memories for different contexts.

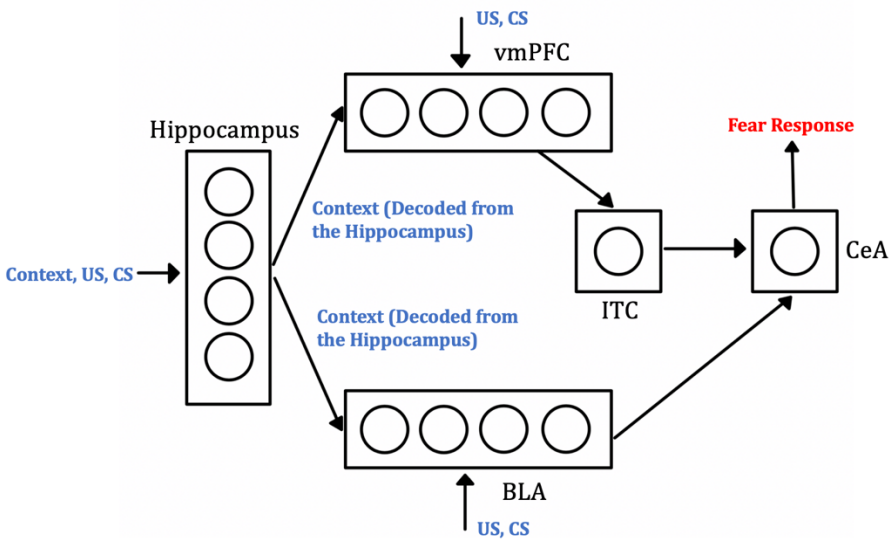
3.4 Ventral Medial Prefrontal Cortex

Since fear extinction involves the formation of new memories rather than erasure of older fear memories, researchers have been looking for the brain region that holds the extinction memories. Studies from mice [7] with ventral medial prefrontal cortex lesions have implied that the vmPFC is what we are looking for. In fact, the output neurons of vmPFC projects to ITC in the amygdala, which is considered as the inhibition site for fear response. This is consistent with the suggestion that the vmPFC is in charge of fear extinction.

4. A Computational Model on Interactions Between the Amygdala, Prefrontal cortex, and the Hippocampus

A computational model for the interactions between the amygdala, prefrontal cortex, and the hippocampus is shown in (Fig.3). The input signals are the conditioned stimulus (CS), unconditioned stimulus (US), and the context. Since the context memories are stored in the hippocampus, the hippocampus decodes the input signals and outputs the encoded context as the form of a sematic pointer, which is sent to both the BLA and vmPFC. Both BLA and vmPFC receives the CS and US as input as well as the decoded context from the hippocampus. Together with the information, the BLA evaluates and outputs a signal encoding whether to excite a fear response, while the vmPFC does the opposite by evaluating whether to inhibit one. Both evaluations are projected to the CeA (the inhibition signal is sent to the CeA indirectly through the ITC), where the two signals compete and the stronger one decides if the

CeA is to initiate a fear response. In a larger model, this fear response signal should be sent into the parasympathetic autonomic nervous system, hypothalamus, and the brainstem motor areas and thus performs the behaviors such as freezing. However, we omit this detail and focus on the fear circuit between the amygdala, prefrontal cortex, and the hippocampus for now.



(Fig. 3) Structure of the Computational Model

4.1 Encoding of the stimuli

There are three stimuli input to the model, a context, CS, and US respectively. Since the input to the model are actually output from sensory systems that projects into the subcortical regions, they may be considered as sematic pointers in the brain. Here we let each stimulus be an 8-bit vector, representing as semantic pointers. The representation of different stimuli or different contexts are orthogonal, as this condition is required in order to ensure that the dissimilarities between all stimuli is the same [8].

4.2 Hippocampus: Model structure and training rules

The hippocampus is a single layer of 20 nodes, each node receives projections from the input signal layer, which is a 24-bit vector that represents the context, CS, and US. The hippocampal layer and input signal layer is fully connected. The weights w_{ji} are initialized uniformly randomly from [0, 0.6] and then perturbed using Gaussian noise (i stands for the index of the nodes in the input layer and j stands for that of the output layer.). Hence the actual weights can be computed as $u_{ji} = w_{ji} + \delta_{ji}$, where δ_{ji} is a sample drawn from a Gaussian distribution with zero mean and variance 0.0025.

The hippocampal layer is trained as a Hebbian Network. Hence, the activation of node j in the hippocampal layer can be computed and updated as the following

equations:

$$y_j(t) = f\left(\sum_{i=1}^{n} u_{ji}(t)x_i(t)\right), \text{where } f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

$$w_{ji}(t+1) = w_{ji}(t) + \alpha_{hipp}x_i(t)y_j(t) \quad (2)$$

The output $\vec{y}(t)$ is sent to both the BLA and vmPFC as the encoded context.

4.3 BLA and vmPFC: Model structure and training rules

Both the BLA and the vmPFC are single layers of 40 nodes as the input is a vector encoding both the context from the hippocampal layer and the original US and CS semantic pointers. The activation method of the units is the same as it is shown in the hippocampal layer in equation (1). However, the training process is different from the hippocampal layer, where the Hebbian Network was used. This is due to the fact that the BLA and vmPFC act like competitors: the BLA is recruited during fear acquisition while the vmPFC is recruited during fear extinction. Research has shown that it is the prediction of the US that decides which of the two part of the brain wins the competition. Hence, the temporal difference (TD-algorithm) algorithm technique [9], a type of reinforcement learning, comes in handy.

The TD-algorithm can be described as follows. If a US is represented but the prediction of the BLA tells the opposite (a positive TD error signal), fear acquisition should be reinforced. On the other hand, if a US is not presented while the prediction of the vmPFC says otherwise (a negative TD signal error), fear extinction should be reinforced. Hence, the BLA and vmPFC should be able to predict the existence of the US based on the trials it has experienced. The TD error can be computed as follows:

$$TD(t) = R(t) + \gamma P(t) - P(t-1) \quad (3)$$

R(t) tells whether is US is actually presented at trial t, which takes on value 1 if the fear US is presented or value 0 if not. γ is the discount factor, which is set to 0.99 in later simulations. P(t) is the prediction made by the BLA and the vmPFC, which can be computed as equation (4).

$$P(t) = \sum_{i=1}^{n} w_i(t)x_i(t) \quad (4)$$

It should be noted that the weights $w_i(t)$ in equation (4) should not be confused with the weights w_{ji} in equation (1) and (2). These two are different weight matrices. The former only exists in the BLA and vmPFC to make predictions on the US in each trial, while the latter are weights that convert input to output in each layer. The weight

updating rules are shown in equation (5)-(8).

The BLA layer is trained on positive TD error signals.

$$w_{ji}(t+1) = w_{ji}(t) + \alpha_{BLA} TD(t) x_i(t) y_j(t) \quad (5)$$

$$w_i(t+1) = w_i(t) + \alpha_{BLA} TD(t) x_i(t) y_j(t) \quad (6)$$

The vmPFC layer is trained on negative TD error signals.

$$w_{ji}(t+1) = w_{ji}(t) - \alpha_{vmPFC} TD(t) x_i(t) y_j(t) \quad (7)$$

$$w_i(t+1) = w_i(t) - \alpha_{vmPFC} TD(t) x_i(t) y_j(t) \quad (8)$$

4.4 Central Nuclei of the Amygdala (CeA): Model Output

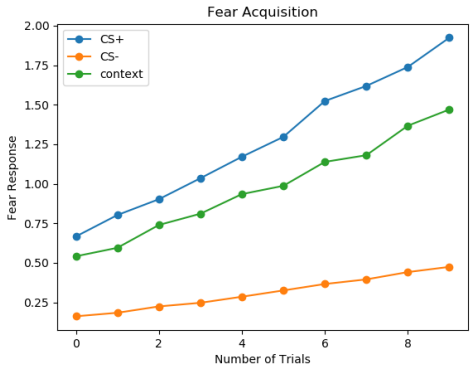
The central nuclei of the amygdala is where fear responses are initiated. It receives direct input from the BLA and indirect input from the vmPFC, which are responsible for fear acquisition and extinction. There is only a node in our model of the CeA, and the output is the strength of the fear response, which is calculated as the difference between the activation of the BLA and vmPFC.

$$CeA(t) = \sum_{j=1}^{n} y_{j,BLA}(t) - \sum_{j=1}^{n} y_{j,vmPFC}(t) \quad (4)$$

5. Results

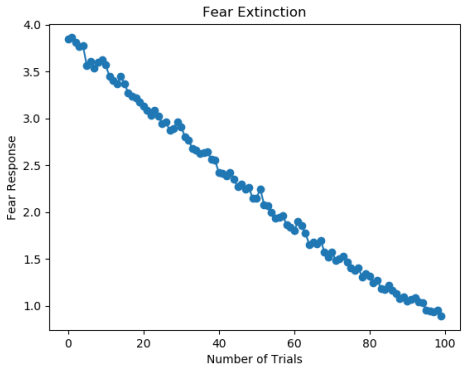
5.1 Fear Acquisition

The fear acquisition process was simulated as pairing of CS and US. 10 trials were simulated, each represents the context, US, CS at the same time. We name the CS that is paired up with US during training process “CS+”. There is also another CS “CS-” but was only presented for testing and absent in the training process. The three lines in the (Fig. 4) show that after a series of trials, the model shows an increased expression of fear response when only CS+ is presented (blue line). This implies that a link was formed between CS+ and US. The context is also able to trigger fear response, though the strength is smaller than the response from CS+. In contrast, the fear response of CS- does not show apparent growth after the trials.



(Fig. 4)

The learning process of fear acquisition



(Fig. 5)

The learning process of fear extinction

5.2 Fear Extinction

Fear extinction can also be simulated by stop presenting the US with the CS. The extinction process is rather slower, where the fear response gradually decreases and eventually falls to the same level before fear acquisition after 100 trials (Fig. 5).

6. Conclusions and Future Work

The fear conditioned circuit is simulated and able to produce results of conditioned fear, including fear acquisition and extinction. Simulation results are consistent with the psychological results in Pavlov conditioning. The computational model proposed not only takes the interactions between the amygdala, the prefrontal cortex, and the hippocampus into account, but it also considers the differentiated nuclei in the amygdala rather than treats it as a black box.

For future work, lesions to different brain regions may be simulated to check if the results are consistent with the biological basis. Another attemptable improvement to the model may be adding other basic emotions aside from “fear”, such as “happy”, “sad”, “angry”, “disgusted”. However, this may be challenging since the biological pathways for those emotions still remains unclear.

7. References

[1] John, Yohan Joshua, et al. "Anatomy and computational modeling of networks underlying cognitive-emotional interaction." Frontiers in human neuroscience 7 (2013): 101.

[2] Bechara, Antoine. "The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage." Brain and cognition 55.1 (2004): 30-40.

[3] Moustafa, Ahmed A., et al. "A model of amygdala–hippocampal–prefrontal interaction in fear conditioning and extinction in animals." Brain and cognition 81.1 (2013): 29-43.

[4] Lotfi, Ehsan, Saeed Setayeshi, and Saeed Taimory. "A neural basis computational model of emotional brain for online visual object recognition." Applied Artificial Intelligence 28.8 (2014): 814-834.

[5] Knight, David C., et al. "Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning." Cognitive, Affective, & Behavioral Neuroscience 4.3 (2004): 317-325.

[6] Anagnostaras, Stephan G., et al. "Scopolamine and Pavlovian fear conditioning in rats: dose-effect analysis." Neuropsychopharmacology 21.6 (1999): 731.

[7] Lebrón, Kelimer, Mohammed R. Milad, and Gregory J. Quirk. "Delayed recall of fear extinction in rats with lesions of ventral medial prefrontal cortex. " Learning & memory 11.5 (2004): 544-548.

[8] Moustafa, Ahmed A., Catherine E. Myers, and Mark A. Gluck. "A neurocomputational model of classical conditioning phenomena: a putative role for the hippocampal region in associative learning." Brain research 1276 (2009): 180-195.

[9] Sutton, Richard S., and Andrew G. Barto. Introduction to reinforcement learning. Vol. 2. No. 4. Cambridge: MIT press, 1998.