Research Proposal

Maria Kesa

The objective of this work is to outline a master's thesis topic. We investigate the emergence of selectivity for positive or negative outcomes in basolateral complex of amygdala (BLA) neurons through training a deep spiking neural network model with a reward-dependent Hebbian plasticity rule, following a recently introduced approach (Engel et al, 2015). Amygdalar circuits are responsible for assigning a “value”-- something worth approaching or avoiding-- in an environment with a changing joint stimulus-reward distribution (Janak and Tye, 2015). Receiving information from the primary sensory cortices and orbital prefrontal cortex (amongst other brain areas not considered in this thesis), the subpopulations in basolateral amygdala send projections to central medial amygdala and nucleus accumbens (amongst other brain areas that are not considered in this thesis), which initiate fearful and reward acquiring behaviors respectively. But how do these cells encode whether the sensory stimulus has value? An extensive line of work has elucidated the changes that take place in the amygdala during Pavlovian fear and reward conditioning. Amongst these results is an intriguing finding that when the cues and the outcomes are exchanged, a significant number of BLA neurons display tuning to the outcomes instead of the sensory cues (Janak and Tye, 2015). Different populations of BLA neurons respond to a cue when the fearful memory persists and after its extinction (Janak and Tye, 2015). Intriguigingly, there is no morphological difference between the principal neurons in BLA. The process of labeling of the stimulus with value is hidden in the connections between the principal neurons and the interneurons in the BLA.

Recent work has revealed an approach for using reward prediction error to train the connections in hierarchical spiking neural networks with sub-networks modeled after biological networks (Engel et al, 2015). The network was trained to discriminate between stimuli, directions in a random moving dots task. The layers of the network consisted of inter-connected sub-networks modeled after biological neurons in MT, LIP and cortical decision making area. The top level of the model emitted choices in response to stimuli according to a winner-take-all dynamics dominated cortical network. The output of this layer was used as a prediction. A novel reward-dependent Hebbian plasticity rule was introduced which formed a prediction error, the difference between the outcome and the expected outcome given the stimulus based on the history of the reward regime and multiplied it with the spiking co-activation between cells. This plasticity rule was used to update the constellation of neural connections between the different layers of the network on every trial. This hierarchical network learned to correctly classify directions 80% of the cases. Furthermore, the top-down prediction error from the decision making circuit resulted in sensory enhancement with category selective cells emerging in LIP as revealed by a Generalized Linear Model analysis. We hypothesize that this same modeling strategy can be fruitfully applied to amygdalar circuits to reveal the mechanisms of the emergence of value selectivity in the BLA.

At the output layer of a hierarchical spiking neural network is a decision layer. In our model formulation we represent decisions about the value of the stimulus as the rate of spiking patterns in nucleus accumbens for appealing stimuli and central medial amygdala for aversive stimuli. BLA in our model is the analogue of LIP in (Engel et al, 2015), serving as an intermediate layer between sensory inputs and decisions. It receives sensory inputs similarly to (1). Prediction errors are used as a feedback mechanism from the decision layer to BLA through Hebbian plasticity and this induces correlations between the decision layer and the BLA layer, which (Engel et al, 2015) calls “choice probability”. It is important to note that in biological brains the backward projections from nucleus accumbens are weak and we incorporate it in the model as a weak connection prior. This is consistent with the Free Energy theory of the brain (Friston,2010), which posits that organisms avoid the thermodynamic fluctuation theorem where the probability of increase in entropy of their states goes up exponentially. They maintain a homeostatic balance that minimizes the entropy of

their states. In this case the weak feedback projections mean that once an organism has learned that approaching an aversive stimulus leads to a bad outcome, this learned bias is relatively difficult to change. The organism learns to avoid the unpredictable stimulus to minimize its bound on surprise, Free Energy. Incidentally, Free Energy optimization is formally equivalent to Hebbian Learning as shown in (Friston, 2010).

I plan to introduce a novelty in the model-- namely introducing learning synaptic plasticities in the BLA via the reward bounded Hebbian algorithm. The original formulation (Engel et al, 2015) did not introduce modifiable synapses within the layers of neurons. The reason is still an experimentally unresolved question of how fear and reward encoding neurons and paravalbuminum and somatostatin interneurons interact amongst themselves (see figure 5b for the possible connectivity profiles from Janak and Tye, 2015. If a network has learned to perform a task the resulting connection strengths have functional signficance and can be analyzed for patterns and because the model is relatively biologically plausible, though simplified, they can lead to experimentally testable predictions.

The models will be implemented in the flexible spiking neural network simulator Brian, which is modular and allows to explicitly write down the equations governing model neurons.

Following the Free Energy discussion, it would be interesting to formalize the hierarchical spiking neural network into a hierarchical Bayesian generative model. If the “value” of the stimulus is represented as a hidden state, then learning, minimizing Free Energy, should lead to an approximate posterior of this value given the neuronal representation of the stimulus, termed the recognition model. There are trade-offs between different approaches to engineering intelligent systems. An approach that tries to model the biology the brain cannot get by without the precise measurements of parameters that govern biological systems. A purely algorithmic approach learns the parameters from data, except for the priors. I believe however that a fruitful approach is to somersault between these two approaches-- as new experimental data arrives, informative models modeling the behavior of the system mechanistically can be constructed and used to increase understanding. Measuring neuronal spikes can be used to elucidate the traces of the algorithms that it implements using information theory (Wibral et al, 2015). Ultimately insights from experimental data can be translated into algorithms as has been the case for Geoffrey Hinton. The inspiration for an algorithm can be derived from anywhere in the universe, from tossing coins in a casino to patterns of rainfall in Africa. Algorithms create their own system and language and completely new modus operandi can emerge. However we still don't understand completely how the brain implements its algorithms and it is therefore fruitful to measure, analyze and build mechanistic models.

I would like to be part of a vibrant research community exploring the principles by which intelligent systems operate and dreaming up novel algorithms for continuously pushing the boundary of what is possible. That is why I wrote this research proposal.

Citations

Engel, T., Chaisangmongkon, W., Freedman, D., Wang, X-J. 2015. “Choice-correlated activity fluctuations underlie learning of neuronal category representation”, *Nature Communications*

Janak, P., Tye, K. 2015. “From circuits to behaviour in the amygdala”, *Nature Reviews*

Friston, K. 2010. “The free-energy principle: a unified brain theory?” *Nature Reviews Neuroscience*

Wibral, M., Lizier, J., Priesemann, V. 2015 “Bits from brains for biologically inspired computing”, *Frontiers in Robotics and AI*