

Thesis title: A neuro-computational model for reward-based motor learning

Candidate : Gianmarco Ragogetti

Tutors: Prof. Angelo Marcelli, Prof. Andrea Viggiano

The following thesis deals with computational models of nervous system employed in motor reinforcement learning. The novel contribution of this work is that it includes a methodology of experiments for evaluating learning rates for human which we compared with the results coming from a computational model we derived from a deep analysis of literature.

Rewards or punishments are particular stimuli able to drive for good or for worse the performance of the action to learn. This happens because they can strengthen or weaken the connections among a combination of sensory input stimuli and a combination of motor activation outputs, attributing them some kind of value.

A reward/ punisher can originate from innate needs(hunger, thirst, etc), coming from hardwired structures in the brain (hypothalamus), yet it could also come from an initially neutral cue (from cortex or sensory inputs) that acquires the ability to produce value after learning(for example money value, approval). We called the formers primary value, while the latter learned values. The efficacy of a stimulus as a reinforcer/punisher depends on the specific context the action take place (Motivating operation).

It is claimed that values drive learning through dopamine firing and that learned values acquire this ability after repetitive pairings with innate primary values, in a Pavlovian classic conditioning paradigm.

Under some hypothesis made we propose a computational model made of:

- A block taking place in Cortex mapping sensory combinations(posterior cortex) and possible actions(motor cortex) . The weights of the net which corresponds to the probability of a movement , given a sensory combination in input. Rewards/punishments alter these probabilities through a selection rule we implemented in Basal Ganglia for action selection;
- A block for the production of values (critic): we evaluated two different scenarios
In the first we considered the block only for innate rewards, made of VTA(Ventral Tegmental Area) and Lateral Hypothalamus(innate rewards) and Lateral Habenula(innate punishments)
In the second scenario we added the structures for learning of rewards, Amygdala, which learns to produce a dopamine activation on the onset of an initially neutral stimulus and a Ventral Striatum, which learns to predict the occurrence of the innate reward, cancelling its dopamine activation.

Innate reward is fundamental for learning value system: even in a well trained system, if the learned stimulus reward is no more able to expect innate stimulus reward(because is occurring late or not at all), and if this occurs frequently it could lose its reinforcing/weakening abilities. This phenomenon is called acquisition extinction and is strictly dependent on the context (motivating operation).

Validation of the model started from Emergent , which provides a biologically accurate model of neuron networks and learning mechanisms and was ported to Matlab , more versatile, in order to prove the ability of system to learn for a specific task .

In this simple task the system has to learn among two possible actions , given a group of stimuli of varying cardinality: 2, 4 and 8.

We evaluated the task in the 2 scenarios described, one with innate rewards and one with learned rewards.

Finally several experiments were performed to evaluate human learning rate: volunteers had to learn to press the right keyboard buttons when visual stimuli appeared on monitor, in order to get an auditory and visual reward.

The experiments were carefully designed in a way such to make comparable the result of simple artificial neural network with those of human performers. The strategy was to select a reduced set of responses and a set of visual stimuli as simple as possible (edges), thus bypassing the problem of a hierarchical complex information representation, by collapsing them in one layer. The results were then fitted with an exponential and a hyperbolic function. Both fittings showed that human learning rate is slow compared to artificial network and decreases with the number of stimuli it has to learn.