Ignasi Cos
Department of Mathematics and Informatics
University of Barcelona

RLMov: Learning the Structure of Movement by Reinforcement

We move to attain rewarding states. However, our understanding of how the central nervous system plans and executes movements remains incomplete. For example, when playing tennis, we may generate reaching movements of remarkable spatial and temporal precision. However, the principles underlying how specific parameters are selected to perform under those constraints are unclear. Which specific principles? Which neural structures are implicated? To study these questions, we performed a set of experiments in which rats were trained to perform a task on a locomotion platform, in which reward delivery was contingent upon spatial and temporal constraints; in other words, the rats had to learn a motor strategy to run against a treadmill to attain a reward to arrive at the designated target position no sooner than at a given time duration. See Rueda-Orozco and Robbe, (2015) for a full description of these experiments. Briefly, their results showed that most rats learned a stereotypical behavioural sequence after some weeks of training. This consisted of an initial dwelling interval, followed by a running against the treadmill interval, to attain the target and the reward (FIG 1). Once the learning phase was over, the rats obtained reward in over 80% of the trials.

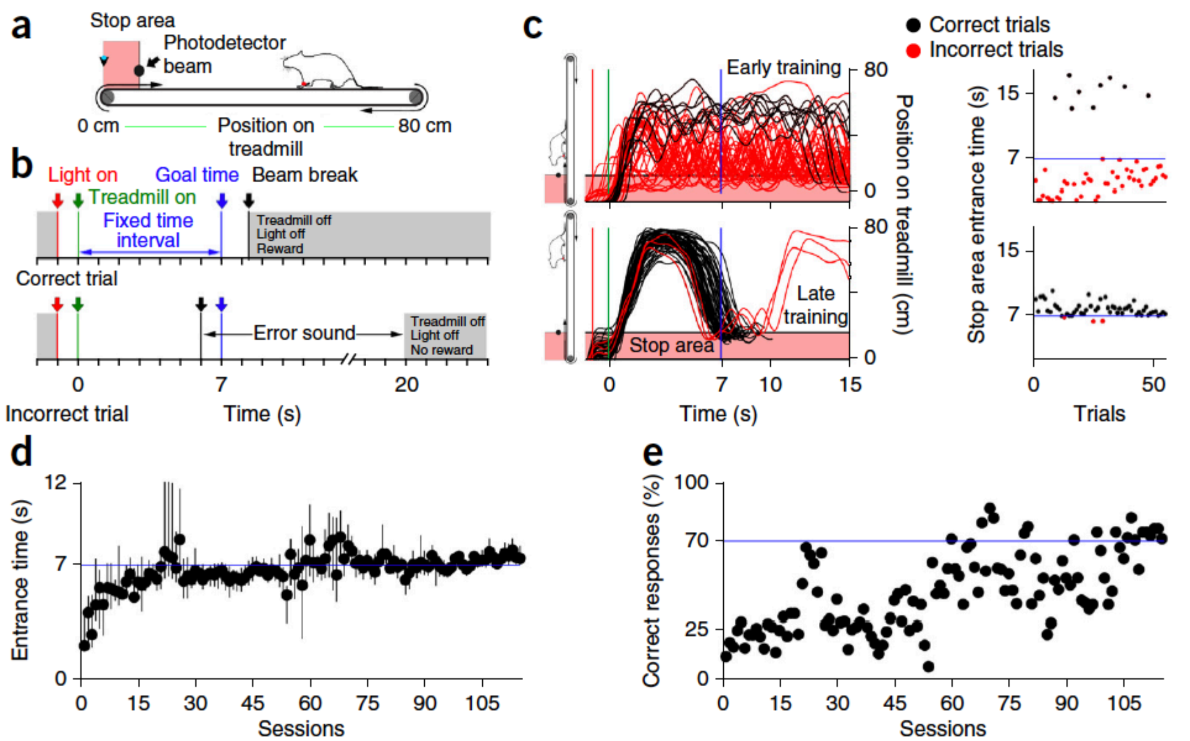

Figure 1 A fixed interval estimation task adapted to treadmill locomotion favored the learning of a stereotyped running sequence. (**a**) Behavioral apparatus. (**b**) Task description. (**c**) Running trajectories (left) and entrance times in the stop area (right) for two sessions, before (top) and after (bottom) learning. (**d**,**e**) Entrance times in the stop area (median ± first and third quartiles; **d**) and percentage of correct trials (**e**) for all the training sessions of a single animal.

Our hypothesis is that the final structure of the rat's behaviour results from a progressively optimized trade-off between the value subjectively perceived when reward is attained, the sensitivity to the effort implied in attaining that reward, and the risk of arriving too early and to miss the reward. We developed some preliminary models of behaviour, implementing the principle of learning sequences to maximize cumulative reward (Cos et al., 2013). Taking this as a starting point, we would be greatly interested in hiring a highly motivated student with a background on Computational Neuroscience. The student must have an interest in reinforcement learning and decision-making, and will have to undertake the design of computational models that could explain the learning and production of these processes within the boundaries of neural motor control and reinforcement learning.

The student's goal would be that of developing reinforcement learning models in the context of motor control and decision-making, to explain the principles underlying the specific organization of behaviour observed experimentally. The student will also be at charge of the preparation of scientific manuscripts for publication and of presenting these results at scientific conferences.

References:

Rueda-Orozco PE, Robbe D (2015) The striatum multiplexes contextual and kinematic information to constrain motor habits execution. Nat. Neurosci. 18:453–460

Cos I, Rueda-Orozco PE, Robbe D, Girard B (2013) Learning Time-Constrained Motor Sequences to Attain Reward: A Speed-Accuracy Trade-off. SfN Abstract.