196. Motivational effects on behavior: Towards a reinforcement learning model of rates of responding

Yael Niv¹, Nathaniel D. Daw², Daphna Joel³, Peter Dayan²
¹ Interdisciplinary Center for Neural Computation, Hebrew University and Gatsby Computational Neuroscience Unit, UCL ² Gatsby Computational Neuroscience Unit, UCL ³ Department of Psychology, Tel Aviv University

There is something rotten in the state of current reinforcement learning models of conditioning. Most experiments report measures such as the rate of responding given the state of the subject and the configuration of the environment. There is even further sub-structure in the temporal organisation of responses, which includes effects associated with the expected timing of forthcoming rewards[1]. However, current theoretical treatments are largely restricted to choices between discrete actions, and are silent about rates or any other aspect of action at a temporal fine scale. This lacuna is brought into sharp relief by current work on the effects of motivation on reinforcement learning, since 'energizing', one of two main motivational effects, is exactly concerned with modulations in the vigour of prepotent responses, and thus changes in rates. Furthermore, dopamine, the golden-haired neurochemical of reinforcement learning, has been directly implicated in these energizing motivational effects[2]. Here, we use data collected from experiments into both energizing, and motivation's other main effect ('directing' behavior, by altering the structure of subjects' goals) to investigate the influence of motivation on free-operant instrumental behavior in detail, in order to propose a reinforcement learning model of rates of responding.

Instrumental behavior has long been associated both with basal-ganglia processing in the brain and with temporal-difference reinforcement learning models. This close relationship between a neural substrate and a computational algorithm stems from the similarity between firing patterns of dopaminergic neurons and the temporal difference prediction-error signal[3], and between the influence of dopamine on plasticity in corticostriatal synapses and learning based on a prediction error in the well known actor-critic model of habitual action control[4]. However, this framework does not allow for the modelling of microscopic effects of motivation and reward expectation on momentary action choice and response rates, as existing reinforcement learning models, stemming from the theory of Markov decision processes, are of discrete choices, rather than rates of behavior.

By conducting a detailed analysis of the momentary behavior of rats lever-pressing for food in either a hungry or a satiated state, we delineate the basic building blocks necessary for a model of on-line control of behavior by motivational states and by reward expectation. In our experiment, levers were 'baited' on a variable-interval schedule. No reward was available for fifteen seconds after a previous reward; thereafter there was a uniform probability for a lever to be baited in the next thirty seconds. The first press on a baited lever would actually produce the sucrose solution reward. We analysed and compared the steady-state behavior of rats extensively trained under this schedule while hungry, to that of rats trained under deprived. Viewing the discrete lever-presses as emitted by a hidden Markov model in which different underlying hidden states generate Poisson-distributed actions with different mean rates, we show that two states (a state of low pressing rate, and a state of high pressing rate) are sufficient to describe the rats' behavior. Further analysis of the individual inter-response-times in search of a molecular structure of bouts of behavior confirmed this conclusion. The overall profile of behavior is that of starting a trial at a low rate of pressing, and transitioning to the higher rate after approximately fifteen seconds, with scalar timing noise evident in this intrinsically timed transition process. Both analyses also show that the hungry rats transition to the high state with a higher probability, and that the rate of lever-pressing in this state is higher in the hungry rats, compared to non-deprived rats.

We aim to incorporate these data into a new reinforcement learning framework which captures the fine-scale structure of behavior. This will take into account the cost and benefit of performing an action, in terms of both the time and effort which the action entails, and its expected value in terms of reward. Actions will be chosen probabilistically based on competition between available actions - some of them instrumental in obtaining experimenter-defined outcomes (such as lever-pressing for food pellets), and some intrinsically rewarding such as exploration, grooming or sleeping. Similar to the derivation of Herrnstein's matching law to single action schedules, such a scheme allows for a hyperbolic relationship between the rate of instrumental behavior and the rate of programmed income. We will simulate behavior on variable-interval schedules of reinforcement and under different motivational states, and compare our results to the experimental data.