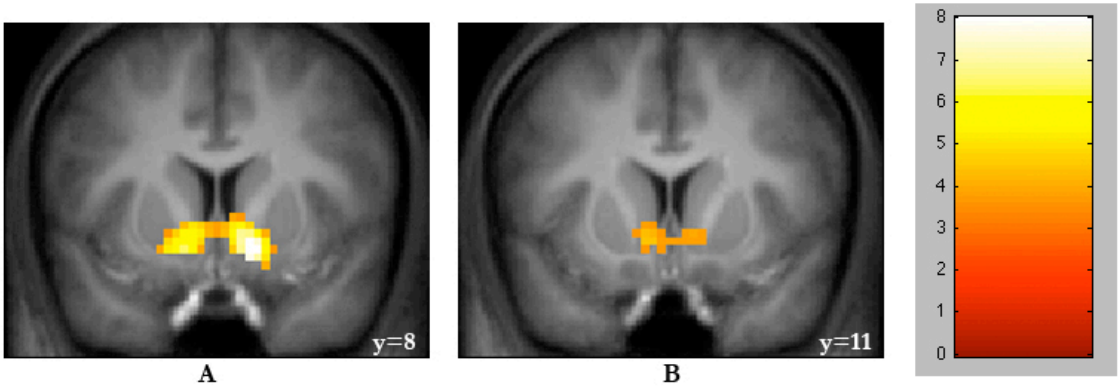
Hierarchical Reinforcement Learning: An fMRI Study of learning in a two-level gambling task

Carlos Diuk, Matthew Botvinick, Andrew Barto and Yael Niv

Hierarchical structure is ubiquitous in human and animal behavior: simple actions are combined to form subtask sequences, which in turn get combined into more complex (and temporally extended) tasks. Here we investigate the neural substrates of hierarchical behavior, based on the computational framework of hierarchical reinforcement learning (HRL). We posit that in order to construct this hierarchical structure, learning must occur at multiple levels at once. The HRL account requires the presence of multiple reward prediction errors, each pertaining to a different level in the hierarchy.

We constructed a two-level "casino task" in which participants were asked to play a hierarchical version of a k-armed bandit task. In each trial participants first chose to play one of two casinos. Once chosen, its door opened and a "house bet" was revealed, along with 4 slot machines. Participants then chose which slot machines to play (for a cost of 2¢ each). To win the casino "house bet" participants needed to win on least two slot machines in the casino. The probability of a win on each slot machine slowly changed throughout the experiment, requiring continuous learning.

This hierarchical design elicited learning at two levels: at the slot machine level for playing within a casino, and at the casino level when having to choose between the two options. Model comparison confirmed that casino-level learning was taking into account both the "house bet" and the overall probability of winning the casino, in true hierarchical fashion. In line with the predictions of HRL, we hypothesized that learning at each level would engender its own prediction error. 21 participants underwent fMRI while playing 100 trials of the casino task. This revealed two distinct BOLD signals correlating with two types of prediction errors: one produced by the outcome of each slot machine play independent of the casino (figure A), and another produced by the outcome of the casino as a whole (figure B). Nucleus accumbens (ventral striatum) activation was significantly correlated with both prediction errors.



A B