Research Notes

Richard Maskell - 1238287

July 11, 2015

1 Uncertainty-Based Competition Between Prefrontal and Dorsolateral Striatal Systems for Behavioural Control

1.1 Document Overview

This article by Daw, Niv & Dayan is based on the idea that "neural systems, notably prefrontal cortex, the striatum and their dopaminergic afferents, are thought to contribute to the selection of actions" (Daw, Niv, and Dayan 2005)

When these systems disagree, the different neurological structures compete with each other, which is modelled here using a "Bayesian principle of arbitration between them according to uncertainty, so each controller is deployed when it should be most accurate". (Daw, Niv, and Dayan 2005)

1.2 The Dorsolateral Striatum

It is stated that "the dorsolateral striatum and its dopaminergic afferents support habitual or reflexive control" (Daw, Niv, and Dayan 2005)

1.3 The Prefrontal Cortex

The prefrontal cortex is said to be "associated with more reflective or cognitive action planning" (Daw, Niv, and Dayan 2005) along with additional regions which are excluded for simplicity in this article.

1.4 Outcome Re-valuation

Conditioning studies where the subject learns desirable behaviour through rewards and/or punishments can have their reward values unexpectedly changed in order to differentiate between two control systems. "Outcome re-valuation affects the two styles of control differently and allows investigation of the characteristics of each controller, its neural substrates and the circumstances under which it dominates" (Daw, Niv, and Dayan 2005)

1.5 Normative Questions

"Why should the brain use multiple action controllers How should action choice be determined when they disagree" (Daw, Niv, and Dayan 2005)

1.6 Reinforcement Learning

"In reinforcement learning, candidate actions are assessed through predictions of their values, defined in terms of the amount of reward they are expected eventually to bring about" (Daw, Niv, and Dayan 2005).

1.6.1 Deferred Rewards

Deferred rewards, such as those dependant on multiple consecutive actions, present complications in predicting the value of actions in reinforcement learning as an initial choice in the sequence may not produce any immediate rewards, only a deferred one. Two different classes of reinforcement learning are used to produce different action's values, which are model-free approaches and model-based approaches. "We interpret the two controllers as representing opposite extremes in a trade-off between the statistically efficient use of experience and computational tractability" (Daw, Niv, and Dayan 2005)

1.6.2 The Model-Free Approach

"[These] underpin existing popular accounts of the activity of dopamine neurons and their (notably dorsolateral) striatal projections" (Daw, Niv, and Dayan 2005). One such method is temporal-difference learning which is founded on the principle of 'caching' which is "the association of an action or situation with a scalar summary of its long-run future value. A hallmark of this is the ubiquitous transfer of the dopaminergic response from rewards to the stimuli that predict them" (Daw, Niv, and Dayan 2005) This method is computationally simple but has the disadvantage of the values being separated from the outcomes and thus, do not change when the outcome is re-valued on the fly.

1.6.3 The Model-Based Approach

"[This approach] involves 'model-based' methods, which we identify with the prefrontal cortex system" (Daw, Niv, and Dayan 2005) The predictions are said to be calculated "on the fly, by chaining together short-term predictions about the immediate consequences of each action in a sequence" (Daw, Niv, and Dayan 2005). Since there is a branching set of situations produced from each action in a state which need to be explored, this is often refered to as 'tree search'. This method can be very computationally expensive due to this exploration of deep trees and can also be erroneous as a result of its complexity. However, an advantage of this approach is that the predicted values are able to react flexibly to outcome re-valuation as they are constructed on the fly.

1.6.4 Accuracy of Different Reinforcement Learning Approaches

The different accuracy ratings of each approach is proposed by Daw, Niv, and Dayan (2005) to justify "the plurality of control and [underpin] arbitration" where the brain relies on the specific control system in the circumstances where it's predictions tend to be most accurate.

Such accuracy is suggested by Daw, Niv, and Dayan (2005) to be estimated for the "purpose of arbitration by tracking the relative uncertainty of the predictions made by each controller."

Strict separation between the systems is assumed in order to isolate their hypothesis

1.7 Results

1.7.1 Post-training Reinforcer Devaluation

Results suggesting circumstances under which each conrtoller dominates

A typical psychological experiments involving post-training reinforcer devaluation is mentioned by Daw, Niv, and Dayan (2005) where "hungry rats are trained to perform a sequence of actions, [...] to obtain a reward such as a food pellet".

1.8 Notes

1.9 Conclusions

References

Daw, Niv, and Dayan (2005). "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioural control". In: *Nature Neuroscience* 8.12, pp. 1704–1711. URL: http://www.nature.com/neuro/journal/v8/n12/abs/nn1560.html.