Psych 202
Dan Birman
Week 7, Law et al. papers

Law et al. lay out in their first paper (Law et al. 2005) an analysis suggesting that the disparate findings in decision making literature can be combined by considering two major sets of findings as evidence for two distinct decision controllers. One controller possibly subsuming a number of pre-frontal regions (but likely overlapping with many other areas) is associated with model-based or goal-directed decision making. They identify the 'devaluation' of a learned stimulus->reward mapping as evidence of model-based learning. In contrast a secondary striatal system is associated with habitual decision making which is not traditionally identified with reward devaluation. They suggest that these two systems can be used in tandem or in isolation depending on task complexity and current ecological demands. To provide evidence for this argument they design a model that can flexibly switch between 'controllers' which implement a TD-learning (model-free) and traditional reinforcement learning (model-based) approaches. They show that for a simple task the model adapts to task demands appropriately, matching behavioral results that they cite. The model also shows the appropriate presence or absence of devaluation which is the behavioral hallmark of each learning system. They conclude that two partially-separable neural systems may be responsible for decision making using vastly different computational implementations. Specifically: during complex or initial learning the pFC may be involved in model-based decision making. Subsequently, for simple and over-learned tasks this learning is transferred to a model-free "habitual" circuitry. In the second paper they substantiate this approach by showing that in certain tasks optimal behavior is a combination of both model-based and model-free learning. They show that in a two-stage decision task participants show both behavior associated with model-based and model-free learning in tandem. Their task involves a first stage decision that probabilistically maps on to two possible second decisions. These second decision stages are then associated with rewards. During the task the probabilistic mappings shift, forcing participants to continuously explore the space to optimize their total reward. They find that striatal BOLD activity associates well with reward prediction error, but the activation is better characterized by a combination of model-based and model-free valuations. This was also true in mPFC. Their overall conclusion is that model-based and model-free reward prediction errors can be usefully integrated. A model-free algorithm trained on the prediction errors of a model-based account would memorize its valuation structure, in theory allowing it to 'habituate' to a task over time.

A question that comes up for me is exactly how the 'switch' to habitual learning occurs over time. As an animal becomes over-trained in a task we don't ever feel a specific switch point where we change our behavior, but nevertheless this evidence suggests that we have a switch (albeit potentially very gradual). It would be interesting to further investigate the 'switch' from model-based to model-free decision making.