How Perceptual Categories Influence Trial and Error Learning in Humans

John McDonnell and Todd M. Gureckis
Department of Psychology
New York University
{*john.mcdonnell,todd.gureckis*}@nyu.edu

Converging evidence suggests that trial-and-error learning in humans shares many computational principals with contemporary RL algorithms (Montague, Hyman, & Cohen, 2004; Frank, Seeberger, & O'Reilly, 2004). However, a critical feature of these algorithms is the notion of "state." State representations help structure which actions to take in particular situations, as well as how to assign credit for delayed rewards. In human reinforcement learning research, the issue of state or context is often avoided by simply assuming that states are interchangeable with the notion of stimulus. Changes in state are simply changes in the currently perceived stimulus. However, the association of states with distinct stimuli is problematic since it limits generalization between different experiences and leads to a combinatorial explosion of potential aspects of the world one must learn about. As a result, human (and artificial) learners must often adopt more nuanced representations that integrate across multiple cues and allow generalization between similar situations. For example, in predicting whether a new restaurant will offer a good meal, many features regarding its decoration, menu, and location may be relevant. These cues combine to allow one to estimate the desirability of the new experience given previous experiences. In addition, a good dining experience at one restaurant may cause changes in the valuation of other, similar restaurants.

Contemporary RL techniques such as function approximation attempt to leverage these ideas in order to support more effective learning. However, less is known about how such "category-like" state representations might influence how people process trial-and-error feedback. We hypothesized that the reinforcement learning system in humans will update valuations of particular state-action pairs with respect to latent classes of stimuli (i.e., clusters or categories) rather than to individual stimuli whenever the stimulus space is too large to memorize all the exemplars. Thus, human learners should behave as though they use a function approximation scheme to structure their learning and should process feedback from different situations contingent on the assignment of particular stimuli to internal stimulus classes.

Overall, our project attempts to leverage fundamental issues concerning state representation in reinforcement learning with recent work in category and concept learning from the psychological literature. One psychological model that has been particularly successful at accounting for human category learning data is SUSTAIN (Love, Medin, & Gureckis, 2004). SUSTAIN assumes that human categories are represented by latent representations call clusters which are recruited to match both the similarity structure of the learning items as well as to reduce overall error. Each category in the model can be represented by one or more clusters which capture salient regularities within the category. In general, the clusters that SUSTAIN recruits depends on the nature of the category and the pattern of feedback the learner receives. Different categories of stimuli can lead the model to adopt different internal representations. To the degree that learners adopt a category representation strategy similar to SUSTAIN, the model predicts that estimates of reward processing may be made with respect to these clustered representations. In particular the reward experienced in response to individual stimuli that belong to the same "cluster" will be averaged together, while reward associated with stimuli belonging to different clusters will be more independent.

To test whether different clustering patterns could lead to different patterns of reward processing, we set up a reward-prediction task we call the 'Mars Weather Prediction' task. On each trial of the task, participants are presented with insect-like creatures (from Mars) and are told to guess what the ratio of two gasses in the atmosphere will be that day under the instruction that the presence of different types of creatures can predict the nature of the martian atmosphere. Participants were rewarded for making accurate estimates, although the actual mixture of gases on any trial was probabilistic (thus perfect accuracy was impossible). The creature stimuli varied on six different features, each of which had a binary value. Examples are 'two legs' vs. 'six legs', or 'round body' vs. 'rectangular body'. Participants indicated their answer by adjusting a continuous slider, and were given corrective feedback.

The task was divided into two phases. In the early training phase, participants were assigned to one of two conditions and were are taught a simple binary categorization over 32 distinct stimuli. The two conditions differed in terms of the representational demands. In the 'family resemblance' (FR) condition, no single feature predicted category membership, and instead items belonging to the same category shared an overall resemblance to one another. In the second condition, the categories were distinguished according to an XOR rule on two pairs of the six stimulus dimensions. SUSTAIN predicts that participants in the FR condition will recruit one cluster for each category (a total of two clusters), while successful learner will recruit four clusters in the XOR condition (representing different components of the XOR rule).

In order to evaluate if such representations actually were used to solve the task, in the second phase, the response distribution of items was shifted. In particular, in the early training phase, there were two response categories: the left side of the slider and the right side. Correct answers on each trial were sampled from gaussian distribution centered at 30% and 70% of the length of the slider (SD=8). In a late training phase, stimuli belonging to the two response categories split in half according to cluster membership, forming response gaussians at 20%, 40%, 60%, and 80% (SD=5.8). The newly split response distributions combined had the same mean and standard deviation as the single distribution they replaced, so, if averaged, they continue to look the same statistically as they did before. If participants are sensitive to the cluster boundaries, they should adapt to the new response structure. If their prior representation tells them to lump the whole category together, they will continue to see each category as having a unitary response structure, 'blind' to the shift that has occurred.

Our results suggest that the structure of cues in the environment strongly influence how reward is updated. In particular, participants appear to show divergent response patterns in based on the types of categories they acquired in the early training phase. FR participants have highly coupled responses to the family resemblance categories and fail to notice changes in the reward structure that are otherwise subsumed by these representations. In contrast, XOR participants tended to decouple the clusters, making them more responsive to the change in reinforcement structure in the later phase of the task. Overall, our results provide insight into how perceptual categories may structure human reward processing. One interesting implication is that participants may show a type of context-dependent TD-error in brain regions associated with error-driven learning such as the basal ganglia. Furthermore, we suggest that the nature of reward processing may reveal aspects of the category representations that human learners adopt.

References

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, *306*, 1940–1943.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychol. Rev.*.

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*, 760–767.