The Impact of Perceptual Aliasing on Human Learning in a Dynamic Decision Making Task

Lisa Zaval, Louis Tur, and Todd M. Gureckis
Department of Psychology
New York University
{*liza.zaval, louis.tur, todd.gureckis*}@nyu.edu

A crucial problem facing both human and artificial RL agents is correctly perceiving, and interpreting, the current state of the environment. For instance, imagine a traveler staying in an unfamiliar hotel, with each floor and exit decorated identically. Based on perceptual information alone, this guest might experience difficulty learning how to navigate towards his room, since the various hallways appear indistinguishable from one another. The problem is one of *perceptual aliasing*, since multiple distinct, task-relevant "states" or situations in the world map to the same percept (Whitehead & Ballard, 1991; McCallum, 1993). Note that environments may be aliased along a continuum from the perspective of the learner (e.g., only every third floor could be identical, or each floor could be perceptually distinct) and may depend on the features that the learner attends to. All else being equal, the ability of the learner should improve as the distortion induced by perceptual aliasing is reduced, and relevant states in the world uniquely map to clearly differentiated percepts.

In this study, we examine how the degree of perceptual aliasing in a task impacts the ability of both human and artificial RL agents to learn effective behavioral strategies in a novel environment. A growing body of work suggests that human trial and error learning shares a similar computational foundation with algorithms developed in the RL literature. However, less work has examined how the identification and categorization of distinct task states might interact with these learning and decision-making processes to determine human performance (see Daw, 2003; Redish, Jensen, Johnson, & Kurth-Nelson, 2007; Veksler, Gray, & Schoelles, 2007 for some related discussion).

Our experiment is based on a dynamic learning and decision-making task called the "Farming on Mars Task" (Gureckis & Love, in press). In the task, participants' make repeated selections between two choices. The goal is to learn a choice allocation strategy that maximizes the total points accumulated over the entire experiment (and participants were paid based on their performance). Unknown to participants at the start of the task, one option appears better than the other on each and every trial because it always results in a larger number of points. However, the current pay-off for either choice is also linked to the relative allocation made to the two options over the last N trials. A dynamic is set up so that each time the more (immediately) attractive alternative is selected, the long-term expected value of both robots is lowered on the following trial. Conversely, selections of the immediately worse option causes the expected value of both options to increase. As a result, the optimal reward-harvesting strategy is to learn to choose the option that appears worse on each individual trial (since it leads to the most long-term reward). The "states" of this task are at best partially observable, representing situations where X of the last N trials are to the immediately attractive option, thus there are N+1 (=10 in our design) task states corresponding to X=0 through X=N.

Previous work with this task found that the performance of human participants can be strongly influenced by the presence of incidental cues which help to differentiate these distinct task states (Gureckis & Love, in press). Gureckis & Love suggested that associating separate perceptual cues with each task "state" help reduce perceptual aliasing and allow more effective learning in the same way that appropriate state representations help artificial learning agents based on Q-learning. However, these previous studies did not directly manipulate the *degree* of aliasing in the task. In the present experiment, we parametrically manipulated the degree of perceptual aliasing given by perceptual cues in the task. In one condition, participants were given no additional cues as part of the display, and thus had to rely on memory and other non-perceptual cues in order to uncover the optimal task strategy. In another condition, the interface screen was augmented with a simple set of cues consisting of two lights. At any point in time, only one of these lights was active, and a shift between the two lights indicated a change in the underlying task system. Importantly, this setup means that 5 distinct task states mapped to the same perceptual display (and thus there was a high degree of aliasing). In a third condition, a circle of five lights was presented on the interface. The indicator lights were organized in a consistent array along the circle, such that the active light moved one position either to the left or right as the state was updated. The five lights were mapped onto the underlying task system using a modulo rule, such that two different task states were mapped to the same perceptual display (thus, the environment was aliased, but less so than in the two-state condition). In the final condition, a display of ten lights was used, such that each active light position corresponded to exactly one task state (no aliasing).

Consistent with Gureckis & Love (2009, in press), our results show that the amount of perceptual aliasing in a sequential decision making task strongly influences participants' ability to learn an optimal response strategy. In particular, reducing perceptual aliasing in the task appears to improve overall task performance. In order to better understand the mechanisms underlying this behavior, we compared the behavioral profile of participant to the prediction of a simple model based on average-reward RL (Gureckis & Love, 2009). This model has previously been shown to effectively predict human choice behavior across a number of similar task scenarios. Interestingly, we find places where the RL model predictions and human behavior diverge. In particular, when participants are given only two state cues (i.e., 5 task states all map to the same percept) performance is actually marginally better than when they are given five cues (2 task states all map to the same percept). This effect was not naturally predicted by the model but suggests a tradeoff between participant's utilization of external state cues (such as perceputal information) and internal memory-based cues (which we model using eligibility traces). Trial-by-trial fits of a slightly modified model which allows for this tradeoff between internal memory and external perceptual cue confirmed these intuitions. Overall, our results show how effective learning in complex tasks depends on a congruence between the way perceptual distinctions in the world relate to task-relevant states. Optimal strategy learning in more difficulty without the aid of stable, perceptual state cues that signal one's present situation but can be compensated for by relying more heavily on memory-based information (McCallum, 1993). In addition to giving insight into the mechanisms underlying human learning, our experimental manipulations and analyses are all motivated by contemporary work in RL algorithms and highlight the potential for constructive dialog between these areas.

References

Daw, N. (2003). *Reinforcement learning models of the dopamine system and their behavioral implications*. Unpublished doctoral dissertation, Carnegie Mellon.

Gureckis, T., & Love, B. C. (2009). Learning in noise: Dynamic decision-making in a variable environment. *Journal of Mathematical Psychology*.

Gureckis, T., & Love, B. C. (in press). Short term gains, long term pains: How cues about state aid learning in dynamic environments. *Cognition*.

McCallum, R. (1993). Overcoming incomplete perception with utile distinction memory. In *The proceedings of the tenth international machine learning conference (ml'93).* Amherst, MA.

Redish, A., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addition, relapse, and problem gamling. *Psychological Review*, *114*(3), 784-805.

Veksler, V., Gray, W., & Schoelles, M. (2007). Categorization and reinforcement learning: State identification in reinforcement learning and network reinforcement learning. In *Proceedings of the 29th annual conference of the cognitive science society.*

Whitehead, S., & Ballard, D. (1991). Learning to perceive and act by trial and error. *Machine Learning*, *7*(1), 45-83.