Brain and Cognition 109 (2016) 1–18

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Dopamine dependence in aggregate feedback learning: A computational cognitive neuroscience approach

Vivian V. Valentin [a,*], W. Todd Maddox [b], F. Gregory Ashby [a]

[a] Department of Psychological & Brain Sciences, University of California, Santa Barbara, United States
[b] Department of Psychology, University of Texas, 108 E. Dean Keeton, Stop A8000, Austin, TX 78712-1043, United States

ARTICLE INFO

Article history:
Received 13 April 2015
Revised 7 June 2016
Accepted 13 June 2016
Available online 3 September 2016

Keywords:
Computational cognitive neuroscience
Dopamine
Skill learning
Striatal plasticity
Parameter space partitioning

ABSTRACT

Procedural learning of skills depends on dopamine-mediated striatal plasticity. Most prior work investigated single stimulus-response procedural learning followed by feedback. However, many skills include several actions that must be performed before feedback is available. A new procedural-learning task is developed in which three independent and successive unsupervised categorization responses receive aggregate feedback indicating either that all three responses were correct, or at least one response was incorrect. Experiment 1 showed superior learning of stimuli in position 3, and that learning in the first two positions was initially compromised, and then recovered. An extensive theoretical analysis that used parameter space partitioning found that a large class of procedural-learning models, which predict propagation of dopamine release from feedback to stimuli, and/or an eligibility trace, fail to fully account for these data. The analysis also suggested that any dopamine released to the second or third stimulus impaired categorization learning in the first and second positions. A second experiment tested and confirmed a novel prediction of this large class of procedural-learning models that if the to-be-learned actions are introduced one-by-one in succession then learning is much better if training begins with the first action (and works forwards) than if it begins with the last action (and works backwards).

1. Introduction

Many skills are acquired via procedural learning, which is characterized by gradual improvements that require extensive practice and immediate feedback (Ashby & Ennis, 2006). Most motor skills fall into this class (Willingham, 1998), but also some cognitive skills, including certain types of category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005, 2010; Maddox & Ashby, 2004). Much evidence suggests that procedural learning is mediated largely within the striatum, and is facilitated by a dopamine (DA) mediated reinforcement learning signal (Badgaiyan, Fischman, & Alpert, 2007; Grafton, Hazeltine, & Ivry, 1995; Jackson & Houghton, 1995; Knopman & Nissen, 1991). The well-accepted theory is that positive feedback that follows successful behaviors increases phasic DA levels in the striatum, which has the effect of strengthening recently active synapses, whereas negative feedback causes DA levels to fall below baseline, which has the effect of weakening recently active synapses (Schultz, 1998). In this way, the DA response to feedback serves as a teaching signal, with successful behaviors increasing in probability and unsuccessful behaviors decreasing in probability.

Experimental studies of DA neuron firing have focused on simple behaviors in which a single cue is followed by a single discrete response (e.g., button or lever press) or no response at all. The seminal finding from these experiments is that DA neurons fire to reward-predicting cues and unexpected reward (e.g. Schultz, 1998). Despite the importance of this work, it does not address the role of DA in the learning of skills that include multiple behaviors that must be precisely executed in response to discrete cues, and in which the feedback is delivered only after the final behavior is complete. Our goal is to investigate the putative role of DA in these more complex settings. We take an indirect approach by collecting behavioral data and then testing a wide variety of computational models that make qualitatively different assumptions about the role of DA in the learning of such multi-step behaviors.

Understanding how multistep behaviors are learned requires an understanding of how the feedback after the final behavior is used to learn the responses to each of the cues in the sequence. One possibility is that feedback propagates backward through each sub-behavior in the sequence, such that the learning of the response to a later cue in the sequence facilitates the learning of a preceding cue. A wealth of data show that once a cue comes to

* Corresponding author.
E-mail addresses: valentin@psych.ucsb.edu (V.V. Valentin), wtoddmaddox@gmail.com (W.T. Maddox), ashby@psych.ucsb.edu (F.G. Ashby).

http://dx.doi.org/10.1016/j.bandc.2016.06.002
0278-2626/

predict reward, it begins to elicit a vigorous response from DA neurons (Pan, Schmidt, Wickens, & Hyland, 2005; Schultz, 1998, 2006; Waelti, Dickinson, & Schultz, 2001). If a new cue is added before a learned cue that perfectly predicts reward, then the DA response to the learned cue shifts back (backpropagates[1]) to the new (earliest) cue (Schultz, Apicella, & Ljungberg, 1993). This works well when no response is required, as in classical conditioning, or in simple instrumental conditioning with only one available response (e.g. lever press), or in tasks requiring choices among different cues while navigating a maze. In such scenarios, DA release due to the reward prediction of the learned cue serves as a teaching signal to train the preceding, new cue, and in this way, sequences of cue-cue associations can be learned (Suri & Schultz, 2001). Importantly, such backpropagation of the DA response has only been demonstrated in tasks in which characteristics of later cues directly depend on decisions made to earlier cues (i.e., *dependent* decisions). Unfortunately, almost no empirical data exist on how DA neurons respond in tasks where a sequence of *independent* decisions must all be made correctly to earn positive feedback, nor have any models been proposed. If several independent decisions about unrelated cues are made in a row, and each has to be correct to earn positive feedback at the end of the sequence, then an earlier cue is not a predictor of a later cue.

Current efforts to study the learning of sequential, multistep decisions have focused on tasks in which the first-step choice predicts the available choices in the next step (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Gläscher, Daw, Dayan, & O'Doherty, 2010; Walsh & Anderson, 2011). This is important work, and many real-life tasks include such dependencies between sequential cues. However, the demonstration in such work that the effect of the feedback backpropagates to earlier cues in the sequence confounds two issues. One possibility is that the backpropagation occurs only because of the perfect dependency, and another is that all sequential skills, including those with independent actions, benefit from such backpropagation. This article investigates the backpropagation of the feedback signal during the learning of a sequence of independent skills. Our results strongly contradict the latter of these two hypotheses. In fact, we show that virtually all models that predict any type of backpropagation of the DA signal to earlier independent cues are incompatible with our results. Furthermore, our results also suggest that any such backpropagation that did occur must have a detrimental effect on learning. Even models that use eligibility traces to update distant cues with the feedback signal (instead of the backpropagation) fail to account fully for our results.

To study how feedback provided at the end of multiple independent behaviors affects the learning of each behavior in the sequence, we developed a new experimental paradigm called the *aggregate-feedback procedural category-learning task* (for a similar declarative memory-based task, see Fu & Anderson, 2008). In this task, three highly discriminable visual images are presented sequentially, each requiring an A or B category response. Feedback is given only after all three responses are complete. Positive feedback is given if all three responses were correct, and negative feedback is given if any of the three responses were incorrect, without any information about which response or responses were in error.

This study addresses a number of fundamental questions regarding DA's involvement in aggregate-feedback learning. These include the following: How do the DA reward prediction signals that develop during learning respond to multiple independent cues before feedback? How does the DA release to the reward prediction of a cue impact learning of cues earlier in the sequence? And do learning rates for cues depend on how far back in time they are from the feedback? We took a *computational cognitive neuroscience* approach to address these questions (Ashby & Hélie, 2011). First, we collected behavioral data from human participants in the aggregate-feedback category-learning task. Second, we used a computational approach called parameter space partitioning (PSP; Pitt, Kim, Navarro, & Myung, 2006) that allowed us to investigate the ability of a broad class of alternative procedural-learning models to account for our results. As we will see, none of these models successfully accounts for all aspects of our data. Third, we used these models to make novel predictions about which of two different training procedures is optimal with aggregate-feedback. Fourth, we tested these predictions with behavioral data from human participants, and identified the best training regime for procedural learning with aggregate feedback.

2. Experiment 1

Our goal was to extend behavioral neuroscience work on DA neuron firing properties to human behavioral experiments. The relevant behavioral neuroscience studies almost all used some form of classical or instrumental conditioning. So the ideal task would share properties with conditioning studies and present some nontrivial cognitive challenges. Our solution was to use an unstructured category-learning task in which highly unique stimuli are randomly assigned to each contrasting category, and thus there is no rule- or similarity-based strategy for determining category membership. This task is similar to instrumental conditioning tasks in which animals must learn to emit one response to one sensory cue and another response to a different cue (e.g., turn left in a T-maze to a high-pitched tone and turn right to a low-pitched tone). But it is also similar to high-level categorization tasks that have been studied for decades in the cognitive psychology literature. For example, Lakoff (1987) famously motivated a whole book on a category in the Australian aboriginal language Dyirbal that includes seemingly unrelated exemplars such as women, fire, dangerous things, some birds that are not dangerous, and the platypus. Similarly, Barsalou (1983) reported evidence that 'ad hoc' categories such as "things to sell at a garage sale" and "things to take on a camping trip" have similar structure and are learned in similar ways to other 'common' categories. Thus, the unstructured category-learning task that forms the foundation of our studies is simple enough that we should be able to relate our results to those from instrumental conditioning studies, while resembling the structure of ad hoc categories.

Although intuition might suggest that unstructured categories are learned via explicit memorization, there is now good evidence – from both behavioral and neuroimaging experiments – that the feedback-based learning of unstructured categories is mediated by procedural memory. First, several neuroimaging studies of unstructured category learning found task-related activation in the striatum, as one would expect from a procedural-learning task, and not in the hippocampus or other medial temporal lobe structures, as would be expected if the task was explicit (Lopez-Paniagua & Seger, 2011; Seger & Cincotta, 2005; Seger, Peterson, Cincotta, Lopez-Paniagua, & Anderson, 2010). Second, Crossley, Madsen, and Ashby (2012) reported behavioral evidence that unstructured category learning is procedural. A hallmark of procedural learning is that it includes a motor component. For example, switching the locations of the response keys interferes with performance in the most widely studied procedural-learning task – namely the serial reaction time task (Willingham, Wells, Farrell, & Stemwedel, 2000). In addition, several studies have shown that switching the response keys interferes with performance of a categorization task known to recruit procedural

[1] Note, our use of the word "backpropagate" refers to the phenomenological dynamics of DA firing to reward predicting events, and not to the popular backpropagation algorithm that is commonly used to train artificial neural networks.