International Journal of Psychophysiology 127 (2018) 62–72

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

# Learning what matters: A neural explanation for the sparsity bias

Cameron D. Hassall[a,*], Patrick C. Connor[b], Thomas P. Trappenberg[b], John J. McDonald[c], Olave E. Krigolson[a]

[a] Centre for Biomedical Research, University of Victoria, Victoria, British Columbia V8W 2Y2, Canada
[b] Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada
[c] Department of Psychology, Simon Fraser University, Vancouver, British Columbia V5A 1S6, Canada

ARTICLE INFO

*Keywords:*
Attention
Learning
Reward positivity
N2pc
Sparsity

ABSTRACT

The visual environment is filled with complex, multi-dimensional objects that vary in their value to an observer's current goals. When faced with multi-dimensional stimuli, humans may rely on biases to learn to select those objects that are most valuable to the task at hand. Here, we show that decision making in a complex task is guided by the sparsity bias: the focusing of attention on a subset of available features. Participants completed a gambling task in which they selected complex stimuli that varied randomly along three dimensions: shape, color, and texture. Each dimension comprised three features (e.g., color: red, green, yellow). Only one dimension was relevant in each block (e.g., color), and a randomly-chosen value ranking determined outcome probabilities (e.g., green > yellow > red). Participants were faster to respond to infrequent probe stimuli that appeared unexpectedly within stimuli that possessed a more valuable feature than to probes appearing within stimuli possessing a less valuable feature. Event-related brain potentials recorded during the task provided a neurophysiological explanation for sparsity as a learning-dependent increase in optimal attentional performance (as measured by the N2pc component of the human event-related potential) and a concomitant learning-dependent decrease in prediction errors (as measured by the feedback-elicited reward positivity). Together, our results suggest that the sparsity bias guides human reinforcement learning in complex environments.

## 1. Introduction

Humans are consistently faced with complex decision problems in multidimensional environments (environments involving choice stimuli made up of numerous features). For example, a decision to eat at one of several restaurants may involve many aspects, including the type of food, ambiance, speed of service, price, availability of parking, location, and existing reviews of the establishment. When combined with other factors (style of cooking, appearance, availability of parking, number of staff, location, reputation, etc.) the state space representing all possible restaurant choices is enormous, yet most humans have little difficulty deciding where to eat. In contrast, reinforcement learning (RL) algorithms that accurately predict behavior in simple tasks become bogged down when faced with multidimensional choices (Sutton and Barto, 1998; also see *curse of dimensionality*: Bellman, 1957). It is therefore problematic that although evidence suggests human learning depends in part on an RL system implemented within dopaminergic midbrain neurons (Roesch et al., 2012; Schultz, 2013) and medial frontal cortex (Holroyd and Coles, 2002; Krigolson et al., 2014; Krigolson et al., 2009; Sambrook and Goslin, 2015), we are able to solve complex multidimensional problems faster than an RL approach alone would predict.

At their core, RL algorithms compare actual feedback with expected feedback to compute prediction errors (Sutton and Barto, 1998), which are then used to update the weights associated with chosen actions. This update process ensures that, in the long run, actions are taken that are more likely to maximize utility (Mill, 1863). Although traditional RL algorithms eventually converge upon optimal solutions, they may do so slowly when compared to an ideal (Bayesian) model. For example, a traditional RL algorithm may select or avoid previously-chosen multidimensional stimuli only insofar as those exact combinations of features are re-encountered. Thus, traditional RL algorithms may be unable to learn about dimensions, only about combinations of features across all dimensions. Consider the following: positive experiences at Italian restaurants in two different areas of town should probably reinforce choosing Italian restaurants in general. However, without rewarding visits to additional Italian restaurants, traditional RL models will only reinforce choosing the two previously visited locations.

The key to the above example is that sometimes only a subset of features is predictive of reward (e.g., style of cooking). Such an

[*] Corresponding author at: School of Exercise Science, Physical and Health Education, University of Victoria, P.O. Box 17000 STN CSC, Victoria, British Columbia V8W 2Y2, Canada.
*E-mail address:* chassall@uvic.ca (C.D. Hassall).

https://doi.org/10.1016/j.ijpsycho.2018.03.006
Received 18 October 2017; Received in revised form 6 March 2018; Accepted 9 March 2018
Available online 15 March 2018

environment is called "sparse" and recent work suggests that humans are able to exploit this property, when it is present (Gershman et al., 2010; Leong et al., 2017; Niv et al., 2015; Wilson and Niv, 2012). For example, Gershman et al. (2010) proposed that humans use an inductive bias – a decision-making assumption – called sparsity to guide learning in multidimensional worlds. The authors presented participants with choice stimuli that varied along three dimensions (shape, color, and texture). Importantly, only one dimension was ever relevant for predicting reward. Gershman et al. (2010) compared human performance on this task with that of a Bayesian model that estimated the posteriors of each feature being the target feature (the likelihoods, given the feedback history). They observed that while humans may not choose optimally (i.e., according to Bayes' rule), they performed better than a traditional RL model (what they called "naive RL") predicted. Based on these findings, Gershman et al. (2010) proposed that humans employ a hybrid RL/Bayesian approach that guides feedback-based learning by selectively attending to a single dimension – the dimension currently believed to be most relevant.

In the present study we assumed that using the sparsity bias would engage two neural processes: an attentional selection process, and an RL process. We therefore hypothesized that a neural marker for each of these processes would be evident in a task for which the use of sparsity was beneficial. We further hypothesized that those neural markers would change with learning as predicted by an attention-weighted RL model (the hybrid RL model proposed by Gershman et al., 2010). We used two event-related potential (ERP) components as neural markers to assess the contributions of both RL and attentional systems within the brain to decision making. First, we measured the N2pc component as an index of selective attention (the proposed mechanism behind the sparsity bias: Gershman et al., 2010; Leong et al., 2017; Niv et al., 2015). Second, we used the reward positivity (Holroyd et al., 2008; Krigolson et al., 2014),[1] also known as the feedback-related negativity (FRN: Miltner et al., 1997), as a neural marker for RL.

A strong relationship between attention and learning is supported by both behavioral and ERP studies. Indeed, our experiences teach us to focus our attention on task-relevant features (Mackintosh, 1975; Pearce and Hall, 1980; Dayan et al., 2000). For example, Wills et al. (2007) showed ERP evidence that unexpected wins (i.e., positive RL prediction errors) grab our attention and drive learning. They did this using stimulus-response learning tasks involving combinations of features. Features that were not expected to elicit a reward (but did) later elicited a large N1, an ERP component associated with the allocation of visual attention to a spatial location (Hillyard and Anllo-Vento, 1998). Eye tracking data confirm that we learn the most about what we attend to; gaze duration for a cue is reduced if it is paired with a previously-rewarded stimulus, or *blocked* (Kruschke et al., 2005). Importantly, attention is not only directed to spatial locations per se, but may also be directed toward features and even feature dimensions (dimensional weighting: Müller et al., 2003).

Our neural marker of selective attention, the N2pc is a negative potential at posterior electrodes that is enhanced contralateral to attended objects appearing in multi-item arrays (Luck and Hillyard, 1994a, 1994b). While it is often observed in response to singletons (single-dimensional stimuli), the N2pc can also be elicited by targets defined by a conjunction of features (e.g., shape and color: Luck et al., 1993). Although there continues to be debate about whether the N2pc reflects enhancement of attended items or suppression of unattended items, researchers agree that the N2pc is tied to an early stage of attentional selection (Eimer, 1996; Hickey et al., 2009; Luck and Hillyard, 1994a, 1994b). Thus, learning biases such as sparsity that rely on selective attention should elicit an N2pc component in response to features that are predictive of reward. Furthermore, we would expect such a component to be dependent on learning. In particular, we might expect an enhanced N2pc component later in learning, when attention is focused on relevant features, compared to early in learning, when the identity of the relevant dimension may be unknown.

While there are currently several ways to measure RL signals in humans (Niv, 2009), a growing body of evidence suggests that the reward positivity indexes a generic RL system within the human brain (Chase et al., 2015; Holroyd and Coles, 2002). The reward positivity component is differentially sensitive to gains and losses (more positive for gains, more negative for losses) and appears 250 to 350 ms after feedback over frontal-central scalp regions. Holroyd and Coles (2002) and others have suggested that the reward positivity reflects an RL prediction error. Specifically, the reward positivity is enhanced for unexpected gains/losses compared to expected gains/losses (Holroyd and Coles, 2002; Holroyd and Krigolson, 2007; Holroyd et al., 2003; Holroyd et al., 2008; Oliveira et al., 2007). Furthermore, the amplitude of the reward positivity tends to decrease with learning as feedback becomes more expected (Krigolson et al., 2014; Krigolson et al., 2009). Finally, the reward positivity appears to shift through time to the earliest indicator that events are better or worse than expected (Holroyd et al., 2011; Krigolson and Holroyd, 2007; Krigolson et al., 2014). (See Niv, 2009, for other ways to measure RL signals in humans.) Thus, two types of reward positivity evidence were considered in the present study. First, the presence (existence) of a reward positivity would indicate the activity of an RL system sensitive to rewards and punishments. Second, changes in the amplitude of the reward positivity ought to behave as predicted by an RL model (e.g., diminish with learning).

The present study tracked changes in two ERP components – the N2pc and the reward positivity – in order to test whether or not humans use the sparsity bias and RL when choosing complex stimuli. If selective attention is engaged, we should observe a (learning dependent) enhancement of the N2pc in response to reward-predicting features, as predicted by a hybrid RL/Bayesian model. If an RL system is engaged, we should observe a reward positivity that decreases with learning, consistent with model-generated prediction errors. We tested these hypotheses by modifying a decision-making task used in previous studies to evaluate the sparsity bias (Gershman et al., 2010; Leong et al., 2017; Niv et al., 2015; Wilson and Niv, 2012). Specifically, we had participants select one of two multidimensional stimuli in which only a single dimension was predictive of reward, and we examined the ERPs elicited by the choice stimuli (for evidence of an attentional bias) and by the subsequent feedback (for evidence of RL), thus providing the first ERP evidence for the sparsity bias.

## 2. Experimental procedures

### 2.1. Participants

We tested 16 participants with no known neurological impairments and with normal or corrected-to-normal vision (4 male, $\mu_{age} = 18.8$ years, $\sigma_{age} = 1.1$ years). All participants provided informed consent approved by the Health Sciences Research Ethics Board at Dalhousie University, and the study was conducted in accordance with the ethical standards prescribed in the original (1964) and subsequent revisions of the Declaration of Helsinki.

### 2.2. Apparatus and procedure

Participants were seated 75 cm in front of a 22-inch LCD monitor (75 Hz, 2 ms response rate, 1680 by 1050 pixels, LG W2242TQ-GF, Seoul, South Korea). Visual stimuli were presented using the Psychophysics Toolbox Extension (Brainard, 1997; Pelli, 1997) for MATLAB (Version 8.2, Mathworks, Natick, USA). Participants were given both verbal and written instructions in which they were asked to minimize head and eye movements.

Participants completed a decision-making task of 40 blocks of 20

[1] For simplicity we will from this point use the term reward positivity. See Proudfit (2015) for a discussion on the definition and naming of the reward positivity/FRN.