Cognition 152 (2016) 160–169

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Neural signature of hierarchically structured expectations predicts clustering and transfer of rule sets in reinforcement learning

Anne Gabrielle Eva Collins [a,b,*], Michael Joshua Frank [a]

[a] Department of Cognitive, Linguistic and Psychological Sciences, Brown Institute for Brain Science, Brown University, Providence, RI, USA
[b] Department of Psychology, University of California Berkeley, Berkeley, CA, USA

ARTICLE INFO

Article history:
Received 19 May 2015
Revised 9 November 2015
Accepted 2 April 2016
Available online 12 April 2016

Keywords:
Structure-learning
EEG
Transfer
Clustering
Prefrontal cortex

ABSTRACT

Often the world is structured such that distinct sensory contexts signify the same abstract rule set. Learning from feedback thus informs us not only about the value of stimulus-action associations but also about which rule set applies. Hierarchical clustering models suggest that learners discover structure in the environment, clustering distinct sensory events into a single latent rule set. Such structure enables a learner to transfer any newly acquired information to other contexts linked to the same rule set, and facilitates re-use of learned knowledge in novel contexts. Here, we show that humans exhibit this transfer, generalization and clustering during learning. Trial-by-trial model-based analysis of EEG signals revealed that subjects' reward expectations incorporated this hierarchical structure; these structured neural signals were predictive of behavioral transfer and clustering. These results further our understanding of how humans learn and generalize flexibly by building abstract, behaviorally relevant representations of the complex, high-dimensional sensory environment.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

How do we take actions that maximize the potential to obtain desired outcomes? Reinforcement-learning (RL) models successfully account for many aspects of human learning behavior and neural activity, by defining a process mechanism that integrates reinforcement history for well-specified stimuli and actions (Frank & O'Reilly, 2006; Montague, Dayan, & Sejnowski, 1996). However, in real life, stimuli are not so well defined: their features are nearly infinite, but only a small subset of them matter for determining how to act. While humans are adept at learning in complex novel situations, RL models in real world settings suffer from the curse of dimensionality. An approach to facilitate learning in complex environments would be to simplify the representation of the environment: for example, to recognize when different sensory states actually should be considered as equivalent, because interaction with them leads to similar outcomes. Doing so would afford generalization and transfer, obviating the need to learn for every single sensory state: given the same goal, any information gathered for one situation may also serve to inform other sensorily distinct, but behaviorally equivalent situations. This "learning to learn" functionality requires building a state and action space that is abstracted away from pure sensory/motor components, but instead comprises functionally relevant states/actions over which RL operates. Computational models of this structure learning process predict that learners cluster together contexts that are indicative of the same latent task set, and further, that such clustering also allows them to construct best guesses of the appropriate set of behaviors in novel contexts (Collins & Frank, 2013). Here, we investigate how the brain constructs, clusters and generalizes these types of structured rule abstractions in the course of learning.

As a real-world example, consider having a laptop with one operating system, and a desktop computer with another. Here, the current sensory context (laptop or desktop) cues a higher-order representation of an abstract context (Mac or Linux), which then determines the lower-order set of rules for behavior (specific actions to reach specific goals). The higher order context defines a rule-set that is "latent" or not tied to a specific context: in this case the observable context is the computer used, but the rule-set is more abstract and can be generalized to other contexts when appropriate, allowing for rapid learning and transfer of new actions. Thus, you may learn that your work desktop is also associated with the "Mac" rule-set. When you learn a new shortcut on that desktop, you can immediately assume that it will have the same effect on your laptop (but not on your home PC) even if you've never tried it before. Similarly, if you try a new computer and the shortcuts typically used on your PC produce desired

* Corresponding author at: 3210 Tolman Hall, University of California Berkeley, Berkeley, CA 94720, USA.
E-mail address: annecollins@berkeley.edu (A.G.E. Collins).

http://dx.doi.org/10.1016/j.cognition.2016.04.002
0010-0277/© 2016 Elsevier B.V. All rights reserved.

effects, you may infer that it has the same OS and generalize your knowledge of that OS to other actions on that new machine. Clustering models further predict that the shortcuts you try in the first place are more likely to be the ones that have worked across a variety of machines – even if they're not the machines (and hence shortcuts) you've used most frequently.

We recently showed that humans build structure *a priori* – subjects do not only discover structure when it exists in the task, but apply structure to learning problems that could be described more simply without structure and in which it is not directly beneficial to learning (Collins & Frank, 2013). Nevertheless, EEG markers of PFC function predicted subjects' tendency to create structure and later use it to generalize previously learned rules to new contexts (Collins, Cavanagh, & Frank, 2014). Computational models captured this structured learning using Bayesian hierarchical clustering (Doshi, 2009; Teh et al., 2006) of task-set rules, which could be approximately implemented in a hierarchical PFC-BG neural network (Collins & Frank, 2013). However, these previous studies were designed to test whether subjects tended to create structure even when no such structure was needed. Here, we develop a paradigm to assess whether subjects discover the form of structure that maximizes their ability to generalize, and whether they do so in a manner predicted by clustering models. In particular, these models predict that subjects should treat a particular dimension of the stimulus to be "higher-order" indicative of the rule-set if distinct elements of that dimension can be clustered together, that is, if they signify the same set of mappings between lower order stimulus features and actions. We test whether subjects can indeed identify the appropriate dimension that affords generalization, and further assess the implications of such clustering in novel contexts. We recorded EEG to assess evidence for such hierarchical clustering in the neural signal.

Specifically, our experimental paradigm (Fig. 1) assesses whether subjects abstract over multiple features that are perceptually distinct (e.g., different colors) but which all signify the same rule in terms of how they condition the contingencies between other features (e.g., shapes), actions and outcomes. Our model predicts that if one feature dimension (e.g. color) allows such clustering of lower level rules, then subjects will treat this feature as higher-order context indicative of an abstract latent task set (Collins & Frank, 2013), while treating the other features (shapes) as lower level stimuli. Because this structure separates the latent rule-set from the contexts (colors) that cue it, it endows a learner with the ability to append any newly encountered lower order stimulus-action associations to an existing rule-set, and thus to immediately generalize it to all contexts indicative of the same set.

Our clustering model makes more specific predictions regarding how subjects treat new contexts in which they are uncertain about which existing rule-set (if any) should apply. Clustering implies not only that contexts indicative of the same rule can be grouped together, but also the number of such contexts in a cluster is indicative of the popularity of that structure, and hence affects the probability that this structure is selected in a new context (technically, we use a non-parametric prior called the Chinese Restaurant Process (CRP) Teh et al., 2006; Gershman & Blei, 2012). Note, however, that the most popular rule may not be the one that has been experienced most often: clustering occurs as a function of number of distinct contexts and not the number of trials (as assumed in other clustering models (Gershman, Blei, & Niv, 2010). (In the computer example, our model predicts that one's expectation for the operating system of a new computer would be based on the relative proportion of computers that had used Mac OS in the subject's experience, even if they had spent 95% of the time on a single PC.) Thus in our design (Figs. 1 and 2A, B) we equate trial frequency across different rule structures but assess whether subjects show evidence of context popularity-based clustering.

EEG is sensitive to reward expectations (Cavanagh, Frank, Klein, & Allen, 2010; Fischer & Ullsperger, 2013; Holroyd & Krigolson, 2007; Holroyd, Pakzad-Vaezi, & Krigolson, 2008; Sambrook & Goslin, 2015; Walsh & Anderson, 2012). We use trial-by-trial model-based analysis (Cavanagh, 2015; Harris, Adolphs, Camerer, & Rangel, 2011; Larsen & O'Doherty, 2014) to investigate whether EEG signals are better accounted for by information processing that includes structure-learning, and whether these signals are predictive of generalization and clustering.

2. Material and methods

2.1. Subjects

2.1.1. Behavioral experiment

34 subjects participated (20 female, ages 18–30), and one was excluded for outlier low performance. Analyses were performed on 33 subjects, including 18 in the TS1 as old TS in phase C group, and 15 in the TS2 group.

2.1.2. EEG experiment

We collected data for 39 subjects (26 female, ages 18–30). 7 subjects were excluded for poor participation (more than 50 no response trials) and a further 3 for poor performance (3 standard deviations under overall group mean performance), so that behavioral analysis was performed on 29 subjects. Due to technical problems with the EEG cap, 3 additional subjects were excluded from EEG analysis, leaving 26 subjects.

2.2. Experimental protocol

2.2.1. Structure

Subjects performed a learning experiment in which they used reinforcement feedback to figure out which key to press for each presented visual input. The experiment was divided into three phases (see Fig. 1C). In all phases, visual input patterns comprised a novel set of colored shapes. After stimulus presentation, subjects selected one of 4 keys to press with their right hand. Simultaneous visual and auditory feedback indicated truthfully whether they had selected the correct action. See Section 2.2.3 for more details.

2.2.2. Phases

The three phases of the experiment were designed to test whether subjects learned hierarchical structure and leveraged it to transfer and generalize knowledge in new contexts. We describe here the protocol in which color acts as "high level" context (Fig. 1A), but the role of color and shape was counterbalanced across subjects. Phase A included 6 different visual stimuli combining one of 3 colors (C0, C1 or C2) and one of two shapes (S1 or S2) (Figs. 1C and 2A, B). We selected colored-shape action associations such that they were identical for C0 and C1 but different for C2. As shown in Fig. 2A, this provides an opportunity for structuring learning such that C0 and C1 can be clustered on a single task-set. Phase B included another 6 different visual stimuli combining one of the same 3 colors (C0, C1 or C2) with one of two new shapes (S3 or S4), Figs. 1 and 2A, B. The associations to be learned in this transfer phase B respected the previous grouping of C0 and C1 into a single task-set (Fig. 2A), such that even though subjects still needed to learn *de novo* the correct actions for the new shapes, we could test whether they could use the structure acquired in phase A to more rapidly learn these associations that are shared between C0 and C1 by generalizing learning from one to the other.

Download English Version:

https://daneshyari.com/en/article/7286181

Download Persian Version:

https://daneshyari.com/article/7286181

Daneshyari.com