Neuroscience Letters 680 (2018) 31–38

Contents lists available at ScienceDirect

Neuroscience Letters

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

Review article

Building concepts one episode at a time: The hippocampus and concept formation

Michael L. Mack[a,*], Bradley C. Love[b,c], Alison R. Preston[d,e,f,*]

[a] Department of Psychology, University of Toronto, Toronto, ON, Canada
[b] Experimental Psychology, University College London, London, UK
[c] Alan Turing Institute, London, UK
[d] Department of Psychology, The University of Texas at Austin, Austin, TX, USA
[e] Center for Learning and Memory, The University of Texas at Austin, Austin, TX, USA
[f] Department of Neuroscience, The University of Texas at Austin, Austin, TX, USA

HIGHLIGHTS

- The hippocampus integrates across experiences to support complex behaviors.
- Activation patterns in the hippocampus are influenced by selective attention.
- These hippocampal processes align with formal accounts of concept learning.
- Recent fMRI evidence supports a role for the hippocampus in concept formation.

ARTICLE INFO

Article history:
Received 18 April 2017
Received in revised form 12 July 2017
Accepted 31 July 2017
Available online 8 August 2017

Keywords:
Hippocampus
Concept learning
Episodic memory
Attention
Prediction error
Computational modeling

ABSTRACT

Concepts organize our experiences and allow for meaningful inferences in novel situations. Acquiring new concepts requires extracting regularities across multiple learning experiences, a process formalized in mathematical models of learning. These models posit a computational framework that has increasingly aligned with the expanding repertoire of functions associated with the hippocampus. Here, we propose the Episodes-to-Concepts (EpCon) theoretical model of hippocampal function in concept learning and review evidence for the hippocampal computations that support concept formation including memory integration, attentional biasing, and memory-based prediction error. We focus on recent studies that have directly assessed the hippocampal role in concept learning with an innovative approach that combines computational modeling and sophisticated neuroimaging measures. Collectively, this work suggests that the hippocampus does much more than encode individual episodes; rather, it adaptively transforms initially-encoded episodic memories into organized conceptual knowledge that drives novel behavior.

Contents

* Corresponding authors.
E-mail addresses: mack@psych.utoronto.ca (M.L. Mack), apreston@utexas.edu (A.R. Preston).

https://doi.org/10.1016/j.neulet.2017.07.061
0304-3940/

1. Introduction

Concepts define the relationships between similar objects; they represent combinations of features shared by objects of the same kind and allow us to recognize new instances of a concept when first encountered. Concepts also serve as the basis for inference about properties that have not or cannot be directly observed. To acquire a concept, we must experience multiple instances across unique episodes and learn both what features are common to concept exemplars and what features differentiate between concepts. Both of these operations, extracting commonalities across related experiences and distinctly representing similar experiences, are akin to episodic memory functions associated with the hippocampus [1–3]. In particular, the hippocampus is thought to perform pattern separation to differentiate overlapping experiences into distinct memory representations [1,2]. Pattern separation is complemented by memory integration, in which the hippocampus is thought to encode features of the current experience along with shared information from previously encoded experiences resulting in integrated memory representations that highlight commonalities across experiences [3,4]. In other words, what concept acquisition requires largely overlaps with coding strategies attributed to the hippocampus.

The theoretical convergence between concept formation and episodic memory posits a role for the hippocampus in acquiring concepts. While initial patient work suggested otherwise [5,6], subsequent findings indicate that the hippocampus plays a key role in representing concepts. For example, “concept cells” in the hippocampus show high selectivity to conceptual rather than perceptual features of events [7] and a recent report showed hippocampal lesions impair concept learning [8]. Here, we review neuroimaging research that has begun to reveal the precise hippocampal mechanisms that support concept formation and use [9–14]. The success of this research has depended on the emergence of sophisticated analytic approaches that combine mathematical accounts of psychological learning theories with representational approaches to neuroimaging. We propose the Episodes-to-Concepts (EpCon) theoretical model of concept formation in the hippocampus, which links evidence from episodic memory and category learning.

2. Building concepts in the hippocampus

It is well established that the hippocampus is critical for rapidly encoding and retrieving experiences to and from memory [15,16]. However, within the past decade, theories of hippocampal function have broadened beyond memory for single episodes [17,18] to suggest that the hippocampus plays the more general role of building flexible representations that span multiple experiences [3], are sensitive to goal states [19,20], and guide novel decisions [21–23]. We propose that this expanded functional repertoire situates the hippocampus as an ideal site for the formation of new conceptual knowledge. Central to this proposal is the EpCon theoretical framework that details how the hippocampus transforms episodic memories to organized concepts.

EpCon is motivated by the striking parallel between hippocampal-based memory processes and a computational model of concept learning named SUSTAIN [24,25]. SUSTAIN posits that during new learning, conceptual representations are formed through a dynamic interaction of selective attention and memory (corresponding hippocampal processes are noted in italics, each of which will be described later):

1) When presented with a stimulus, attention is directed to stimulus feature dimensions that are diagnostic for the task goal according to the current state of knowledge (*attentional biasing*).
2) The attention-weighted feature information then promotes retrieval of similar prior learning experiences (*pattern completion*). These memories are used to predict a concept label.
3) Depending on the prediction outcome (*memory-based prediction error*), a new distinct memory is created that binds together the current stimulus and the correct concept (*pattern separation*), and/or an existing concept representation is updated to incorporate the new stimulus (*integration*). This updated knowledge state then influences attentional strategy on subsequent learning experiences.

As learning continues, this process iterates. Pattern completion retrieves previously integrated representations that highlight the common features diagnostic of the concept, which, in turn are updated with new information from the current experience. Irrelevant features are dropped from concept representations, and concept exemplars are organized according to their similarity on the most relevant features, with the most typical exemplars taking a central position in representational space. By learning what features are common to concept exemplars and what features differentiate between concepts, this adaptive process transforms initially-encoded episodic memories into organized conceptual knowledge representations (Fig. 1).

The component processes of this theoretical framework for concept learning map onto the hippocampal functions of pattern separation and completion, memory integration, and memory-based prediction error, and the framework is further influenced by the fact that hippocampal encoding is biased by attention. The EpCon model is thus a theoretical bridge between SUSTAIN's formalism of concept learning and the functions of the hippocampus. It is important to note that concept learning is supported by many brain regions (see [26] for a recent review); EpCon serves to highlight how the hippocampus is an important player in concept learning's broader neural substrate. Below, we review the evidence for EpCon by highlighting the complementary hippocampal functions that are implicated in concept formation.

2.1. Memory integration

Memory integration arises when the current experience shares features with previously-encoded experiences, which may trigger hippocampal pattern completion resulting in the retrieval of related memories. The current experience may then be encoded into the reactivated memory trace, resulting in an updated representation that captures both the features of the current experience as well as those of the retrieved memory [1,3,27,28]. A wave of recent findings has converged on the existence of such integrated representations in the hippocampus that support complex inference behaviors [29–34].

In particular, one recent human fMRI study by Schlichting et al. [32] targeted the specific nature of integrated representations in the hippocampus. In this study, participants learned pairs of novel