

Journal Club

Editor's Note: These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at www.jneurosci.org. For more information on the format, review process, and purpose of Journal Club articles, please see <http://jneurosci.org/content/preparing-manuscript#journalclub>.

A Novel Role for the Hippocampus in Category Learning

 Sarah A. Hutter and  Andrew I. Wilson

Princeton University, Department of Psychology, Princeton Neuroscience Institute, Princeton, New Jersey 08544
Review of Bowman and Zeithamova

The hippocampus has long been thought to play an important role in remembering details of specific experiences. Only recently has it been implicated in integrating information across events to generalize knowledge (Preston et al., 2004; Zeithamova et al., 2012; Pakkert et al., 2017). One such type of generalization, episodic inference, allows memory of past and current events to be combined so that one can make inferences requiring knowledge from both episodes. Studies of episodic inference have defined a role for the hippocampus in conjunction with the ventromedial prefrontal cortex (VMPFC) in generalization (Schlichting and Preston, 2015). However, whether VMPFC and hippocampal networks contribute to other kinds of generalization, such as concept learning, has been unclear.

Concept learning requires defining new categories based on trial and error. Research on patients with hippocampal damage suggests that concept generalization does not depend on the hippocampus (Knowlton and Squire, 1993; Filoteo et al., 2001), but categorization in healthy individuals is successfully tracked by hippocampal activation (Zeithamova et al., 2008). Still, the precise role of this activity

is debated. Some have argued that the hippocampus retrieves exemplars, previously learned members of a category, to compare against current, novel stimuli (Koenig et al., 2008). Others have found evidence that the hippocampus integrates information across multiple exemplars (Collin et al., 2015; Schlichting and Preston, 2015). Such integration might be the basis of a prototype strategy, in which the attributes of known members of the category are generalized to form an “average member” that is later accessed for comparison with novel stimuli. There are important theoretical distinctions between comparing novel stimuli with exemplars and comparing with a prototype (Zaki et al., 2003). However, it is unknown which, if either, of these strategies occurs in the hippocampus.

Bowman and Zeithamova (2018) sought to determine whether the hippocampus consistently uses either an exemplar or prototype strategy by combining model-based fMRI analysis with a conceptual generalization task. The behavioral paradigm, borrowed from Bozoki et al. (2006), has participants categorize cartoon animal stimuli that differ along eight dimensions (e.g., foot shape, head orientation, color), each of which could assume 1 of 2 values. A particular stimulus was randomly designated as the prototype for the first category, and the “opposite” stimulus (along all dimensions) served as the prototype for the other category. These prototypes define a linear space of stimuli with each prototype at one extreme, allowing every other stimulus to be numerically ranked according to

its perceptual distance from the prototypes. Stimuli are therefore assigned to the category with the prototype for which they have greater perceptual similarity. Participants were trained, with feedback, to assign a subset of the available stimuli to the appropriate category. Importantly, the two category prototypes were not used in the training set. In a generalization phase following training, participants were shown the full spectrum of stimuli, including category prototypes, and were asked to categorize them without feedback. Participants underwent fMRI during this generalization phase.

To distinguish whether participants used prototype and exemplar strategies, the authors fit two previously established models to the trial-by-trial behavioral data (Nosofsky, 1987; Minda and Smith, 2001; Zaki et al., 2003; Maddox et al., 2011). The prototype model considers the expected response of participants, given the distance of the current stimulus from the category prototype. The exemplar model considers the expected response of participants given the perceptual distance of the current stimulus from known members of the category (Bowman and Zeithamova, 2018, their Fig. 1*D*). While both models performed better than chance, the prototype model provided a better fit to the data in 73% of participants. This suggests that most participants compared stimuli with stored prototypes (i.e., integrated evidence from previous experiences) as opposed to exemplars.

Received April 28, 2018; revised June 25, 2018; accepted June 25, 2018.

We thank Jordan Taylor and Michael Graziano for advice and suggestions.

The authors declare no competing financial interests.

Correspondence should be addressed to Sarah A. Hutter, 401 Peretsman-Scully Hall, Department of Psychology, Princeton University, Princeton, NJ 08544. E-mail: shutter@princeton.edu.

DOI:10.1523/JNEUROSCI.1085-18.2018

Copyright © 2018 the authors 0270-6474/18/386803-03\$15.00/0

To investigate the neural contributions to this process, Bowman and Zeithamova (2018) conducted model-based analyses on specific ROIs and, separately, over the whole brain. They found that activity in VMPFC and anterior hippocampus tracked category membership probability outputs from the prototype model but not the exemplar model. That is, an increase in membership probability for a given category is correlated with an increase in category-specific hippocampal activity. Whole-brain analysis further identified several regions tracking outputs from the prototype model. The dominance of prototype-model correlates in the brain activity nicely corroborates the behavioral results.

A critical aspect of using prototypes for categorization is that it requires the brain to construct and store those prototypes in long-term memory for later comparison with novel stimuli. Those prototypes must be constantly updated during training to accurately reflect the central tendency of the evidence gathered for the relevant categories. That integrated category information must then be reinstated and referenced during generalization. Bowman and Zeithamova (2018) suggest that integrated information is represented in VMPFC and the anterior hippocampus, given the presence of prototype-correlated activity, but not exemplar-correlated activity (which does not require the construction of abstract concepts), in these areas during generalization. No exemplar correlates emerged in the initial analyses; however, when significance thresholds were set more leniently, activity correlating with category membership predicted by an exemplar strategy emerged in several areas, such as lateral occipital and inferior parietal cortex. It must be noted, however, that exemplar-correlated activity would not be expected in any case, given the behavioral results.

The authors conclude that a participant using a prototype-based categorization strategy calculates the distance of a presented stimulus from stored representations in VMPFC and anterior hippocampus. At a neural level, this could be achieved by comparing sustained firing patterns, representing a prototype stimulus, with those evoked by the present stimulus. The calculated distance information is then entered as evidence for or against the stimulus belonging to each category during the decision-making process. However, it is unclear whether this is the only comparison being made in VMPFC and hippocampus. It is possible, given the

presented results, that participants simultaneously considered outputs from an exemplar-based categorization process. In both behavioral and fMRI analyses, a weak signal was present for exemplar use and a strong signal for prototype use. The authors suggest that both strategies may be used in the brain, but that their particular task encouraged a prototype strategy.

Supporting the above conclusion, other work has identified exemplar correlates under different task conditions (Mack et al., 2013). This raises the question of how the representation of one strategy comes to dominate over the other, specifically in terms of the strength of that representation and its resultant behavioral output. It is possible that evidence for the efficacy of each categorization model is accumulated over the course of training. Accumulation may occur when the models give differing estimates of category membership probability during training. In each instance of feedback, the more accurate model, in terms of model fit error, could be strengthened and the other weakened. The decision “weight” of each categorization strategy could be tracked by the associated overall network activity (e.g., VMPFC and anterior hippocampus for the prototype model). The network with greater activity, based on the weights as accumulated in training, would then be reflected in behavior, as seen in Bowman and Zeithamova (2018). In other words, the appropriate strategy would come to more strongly influence the categorization decision process over the course of training. When prototype and exemplar models give differing outputs for a given stimulus, this process would serve to increase decision certainty by biasing the decision-making process toward a model that yields better results.

The above account suggests a theoretical mechanism for determining the relative influence of exemplar or prototype strategies during categorization. It serves to reconcile the prototype-based evidence accumulation described by Bowman and Zeithamova (2018) with other findings that describe category determination by comparison with exemplars (Mack et al., 2013). Our proposed model for category learning tasks is as follows. Sensory information is received and fed into both exemplar and prototype models and the evidence is weighted according to the relative network activity associated with each model. The weighted evidence is passed to an accumulator, which collects evidence until a threshold is reached, at which

point a decision is made (as in the drift-diffusion model) (for review, see Bogacz et al., 2006). Accumulators are populations of neurons, previously observed in dorsolateral PFC (van Veen et al., 2008), that increase in activity as evidence in support of a conclusion is gathered. In this case, evidence is accumulated to support one model over the others. The output of the accumulator, a decision, results in behavior. The subsequent feedback resultant from that behavior is then used to reweight the model-related network activity. This process allows convergence to the category model that best predicts reward. By this account, Bowman and Zeithamova (2018) find evidence for prototype correlates in the brain because their task rewards prototype-driven behavior, which in turn increases the weights on prototype network activity. In contrast, Mack et al. (2013) observe exemplar correlates because reward contingencies in their task favor exemplar-driven behavior.

The model we propose bridges the gap between the neural and behavioral results observed in Bowman and Zeithamova (2018). Fully accounting for behavior, from perception to action, opens the door to specific, theory-driven hypotheses and complete computational accounts of how neural processing drives behavior.

References

- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113:700–765. [CrossRef Medline](#)
- Bowman CR, Zeithamova D (2018) Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *J Neurosci* 38:2605–2614. [CrossRef Medline](#)
- Bozoki A, Grossman M, Smith EE (2006) Can patients with Alzheimer’s disease learn a category implicitly? *Neuropsychologia* 44:816–827. [CrossRef Medline](#)
- Collin SH, Milivojevic B, Doeller CF (2015) Memory hierarchies map onto the hippocampal long axis in humans. *Nat Neurosci* 18:1562–1564. [CrossRef Medline](#)
- Filoteo JV, Maddox WT, Davis JD (2001) Quantitative modeling of category learning in amnesic patients. *J Int Neuropsychol Soc* 7:1–19. [CrossRef Medline](#)
- Knowlton BJ, Squire LR (1993) The learning of categories: parallel brain systems for item memory and category knowledge. *Science* 262:1747–1749. [CrossRef Medline](#)
- Koenig P, Smith EE, Troiani V, Anderson C, Moore P, Grossman M (2008) Medial temporal lobe involvement in an implicit memory task: evidence of collaborating implicit and explicit memory systems from fMRI and Alzheimer’s disease. *Cereb Cortex* 18:2831–2843. [CrossRef Medline](#)

- Mack ML, Preston AR, Love BC (2013) Decoding the brain's algorithm for categorization from its neural implementation. *Curr Biol* 23:2023–2027. [CrossRef Medline](#)
- Maddox WT, Glass BD, Zeithamova D, Savarie ZR, Bowen C, Matthews MD, Schnyer DM (2011) The effects of sleep deprivation on dissociable prototype learning systems. *Sleep* 34:253–260. [CrossRef Medline](#)
- Minda JP & Smith JD (2001) Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:775–799. [CrossRef Medline](#)
- Nosofsky RM (1987) Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:87–108. [CrossRef Medline](#)
- Pajkert A, Finke C, Shing YL, Hoffmann M, Sommer W, Heekeren HR, Ploner CJ (2017) Memory integration in humans with hippocampal lesions. *Hippocampus* 27:1230–1238. [CrossRef Medline](#)
- Preston AR, Shrager Y, Dudukovic NM, Gabrieli JD (2004) Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus* 14:148–152. [CrossRef Medline](#)
- Schlichting ML, Preston AR (2015) Memory integration: neural mechanisms and implications for behavior. *Curr Opin Behav Sci* 1:1–8. [CrossRef Medline](#)
- van Veen V, Krug MK, Carter CS (2008) The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *J Cogn Neurosci* 20:1952–1965. [CrossRef Medline](#)
- Zaki SR, Nosofsky RM, Stanton RD, Cohen AL (2003) Prototype and exemplar accounts of category learning and attentional allocation: a reassessment. *J Exp Psychol Learn Mem Cogn* 29:1160–1173. [CrossRef Medline](#)
- Zeithamova D, Maddox WT, Schnyer DM (2008) Dissociable prototype learning systems: evidence from brain imaging and behavior. *J Neurosci* 28:13194–13201. [CrossRef Medline](#)
- Zeithamova D, Dominick AL, Preston AR (2012) Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron* 75:168–179. [CrossRef Medline](#)