Published in final edited form as:
Trends Cogn Sci. 2014 February ; 18(2): 67–69. doi:10.1016/j.tics.2013.10.014.

An Exemplar of Model-Based Cognitive Neuroscience

Thomas J. Palmeri
Vanderbilt University

Abstract

Are categories learned by forming abstract prototypes or by remembering specific exemplars? Mack, Preston, and Love observed that patterns of fMRI brain activity were more consistent with patterns of representations predicted by exemplar models than prototype models. Their work represents the theoretical power of emerging approaches to model-based cognitive neuroscience.

A primary aim of cognitive science is to understand the mechanisms that give rise to faculties of mind like perception, learning, and decision making. One approach formalizes hypotheses about cognitive mechanisms in computational models. Cognitive models predict behavior, like the errors people make and the time it takes them to respond, and how behavior varies under different conditions, using different stimuli, with different amounts of learning. Another approach turns to the brain to identify neural mechanisms associated with different aspects of cognition, using techniques like neurophysiology, electrophysiology, and functional magnetic resonance imaging (fMRI).

These two come together in a powerful new approach called model-based cognitive neuroscience [1]. Cognitive models decompose complex behavior into representations and processes and these latent model states are used to explain the modulation of brain states under different experimental conditions. Reciprocally, neural measures provide additional data that help constrain cognitive models and adjudicate between competing cognitive models that make similar predictions of behavior. For example, brain measures are related to cognitive model parameters fitted to individual participant data [2], measures of brain dynamics are related to measures of model dynamics [3-4], model parameters are constrained by neural measures [4], model parameters are used in statistical analyses of neural data [5], or neural data, behavioral data, and cognitive models are analyzed jointly within hierarchical statistical framework [6].

Mack, Love, and Preston [7] adopted a model-based cognitive neuroscience approach to understand the mechanisms involved in category learning [8]. Consider everyday categories like *dogs*, *cars*, or *chairs*. Categories like these are abstractions in the sense that collections of visibly different objects are treated as the same kind of thing. But does that imply that the mental representations of categories are inherently abstract and that category learning involves creating abstractions? The earliest work on categorization assumed abstraction, either in the form of logical rules defining category membership, or in the form of abstract prototypes capturing the family resemblance of category members. However, later work

address correspondences to: Thomas J. Palmeri 301 Wilson Hall Department of Psychology Vanderbilt University Nashville, TN 37240 tel: 615-343-7900 fax: 615-343-8449 thomas.j.palmeri@vanderbilt.edu web: catlab.psy.vanderbilt.edu.

showed that cognitive models based on memory for experienced category exemplars could predict experimental results that seemed to instead suggest abstraction. While many argue that the evidence favors exemplar models, debate about exemplar models versus prototype models continues [8-10]. Could patterns of brain activity help adjudicate this theoretical controversy?

In [7], prior to scanning, participants learned to classify novel objects into one of two categories. Using a standard category learning procedure [8], over several training blocks, participants viewed an object on each trial, categorized it as a member of Category A or B, and received corrective feedback. In the scanner, participants categorized training objects and new transfer objects as members of Category A or B without feedback (Fig. 1a).

Mack and colleagues [7] used common mathematical formalizations of exemplar and prototype models, fitting them to the probability of categorizing objects as a member of each category for every participant (Fig. 1a). The models make the same assumptions about how objects are represented, how similarities between objects and stored representations are computed, and how categorization decisions are made. Both models assume that categorization decisions are based on the relative similarity of an object to stored category representations. Naturally, they differ in the nature of those representations. For the exemplar model the evidence that an object is a member of Category A is based on the summed similarity of the object to stored exemplars of Category A divided by the summed similarity to stored exemplars of both categories, while for the prototype model the evidence is based on the similarity of the object to the prototype of Category A divided by the summed similarity to prototypes of both categories.

The summed similarity to the stored category representations – summed similarity to exemplars for the exemplar model versus summed similarity to prototypes for the prototype model – constitutes a latent model signature that Mack and colleagues called *representational match*. Although when fitted to behavioral data, the exemplar and prototype models make similar quantitative predictions about the probability that any given object is categorized as an A or a B, they differ considerably in the representational match for any given object which governs its predicted categorization (Fig. 1b). Are the patterns of brain activity measured by fMRI while participants categorize each object more consistent with the representational match predicted by an exemplar model or a prototype model?

It is common to use multivoxel pattern analysis (MVPA) to identify patterns of brain activity that predict different kinds of stimuli, responses, or conditions. In [7], the goal was instead to use MVPA to identify patterns of brain activity that predict different values of representational match for different objects, where values of representational match came either from fits of the exemplar model or the prototype model to individual participant categorization data. A mutual information (MI) measure was used to quantify the relationship between brain states and latent model states, with higher MI reflecting greater consistency between patterns of voxel activity in the brain and patterns of representational match predicted by a model. The exemplar model was more consistent with brain measures than the prototype model, producing significantly greater MI measures (Fig. 1c,d).

In [7], the exemplar and prototype models make nearly identical predictions about behavior. So comparing patterns of brain states with patterns of behavior, as might be traditionally done in cognitive neuroscience, would never uncover how the brain represents categories. Instead, by comparing how patterns of brain states compare with predicted latent model states we can begin to answer to this fundamental question. Categories are learned by remembering exemplars not abstracting prototypes [2, 8-9]. With its joint use of computational models of cognition with brain measures, this work well illustrates the

growing sophistication and theoretical power of model-based cognitive neuroscience approaches [1-6].

Acknowledgments

This work was supported by NSF grant SMA 1041755 (Temporal Dynamics of Learning Center) and NEI grant R01 EY21833.

References

[1]. Forstmann BU, et al. Reciprocal relations between cognitive neuroscience and cognitive models: Opposites attract? Trends Cog. Sci. 2011; 6:272–279.

[2]. Nosofsky, et al. Activation in the neural network responsible for categorization and recognition reflects parameter changes. Proc. Natl. Acad. Sci. USA. 2012; 109:333–338. [PubMed: 22184233]

[3]. Davis T, et al. Learning the exception to the rule: model-based FMRI reveals specialized representations for surprising category members. Cereb. Cortex. 2012; 22:260–273. [PubMed: 21666132]

[4]. Purcell BA, et al. From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search. J. Neurosci. 2012; 32(10):3433–3446. [PubMed: 22399766]

[5]. White CN, et al. Perceptual criteria in the human brain. J. Neurosci. 2012; 32(47):16716–16724. [PubMed: 23175825]

[6]. Turner BM, et al. A Bayesian framework for simultaneously modeling neural and behavioral data. Neuroimage. 2013; 72:193–206. [PubMed: 23370060]

[7]. Mack ML, et al. Decoding the brain's algorithm for categorization from its neural implementation. Curr. Biol. 2013 (2013), DOI: 10.1016/j.cub.2013.08.035 (http://www.cell.com/current-biology).

[8]. Richler JJ, Palmeri TJ. Visual category learning. Wiley Interdiscip. Rev. Cogn. Sci. (in press).

[9]. Nosofsky RM. Exemplar representation without generalization? Comment on Smith and Minda's (2000) Thirty categorization results in search of a model. J. Exp. Psychol. Learn. Mem. Cogn. 2000; 26(6):1735–1743. [PubMed: 11185793]

[10]. Smith DJ, Minda JP. Thirty categorization results in search of a model. J. Exp. Psychol. Learn. Mem. Cogn. 2000; 26(1):3. [PubMed: 10682288]

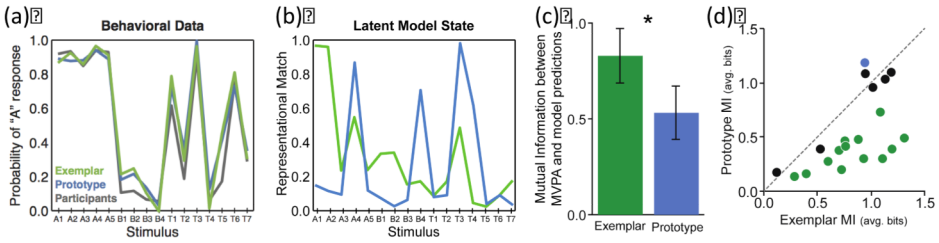

Figure 1.

(a) Probability of a Category A response for each training stimulus (A1-A5, B1-B4) and transfer stimulus (T1-T7) observed from participants (gray), predicted by the exemplar model (green), and predicted by the prototype model (blue). **(b)** Latent model state (representational match) for each training stimulus (A1-A5, B1-B4) and transfer stimulus (T1-T7) predicted by the exemplar model (green) and prototype model (blue). **(c)** Correspondence (mutual information, MI) between patterns of brain activity revealed by multivariate pattern analysis (MVPA) and representational match predicted by the exemplar model (green) and prototype model (blue); higher MI means closer correspondence. **(d)** Mutual information (MI) between MVPA and model predictions (representational match) for individual participants; correspondence for thirteen participants was significantly better for the exemplar than the prototype model (green), for six there was no significant difference (black), and for only one participant it better for the prototype than the exemplar model (blue). Adapted with permission from [7].