Tracking item representations during free recall

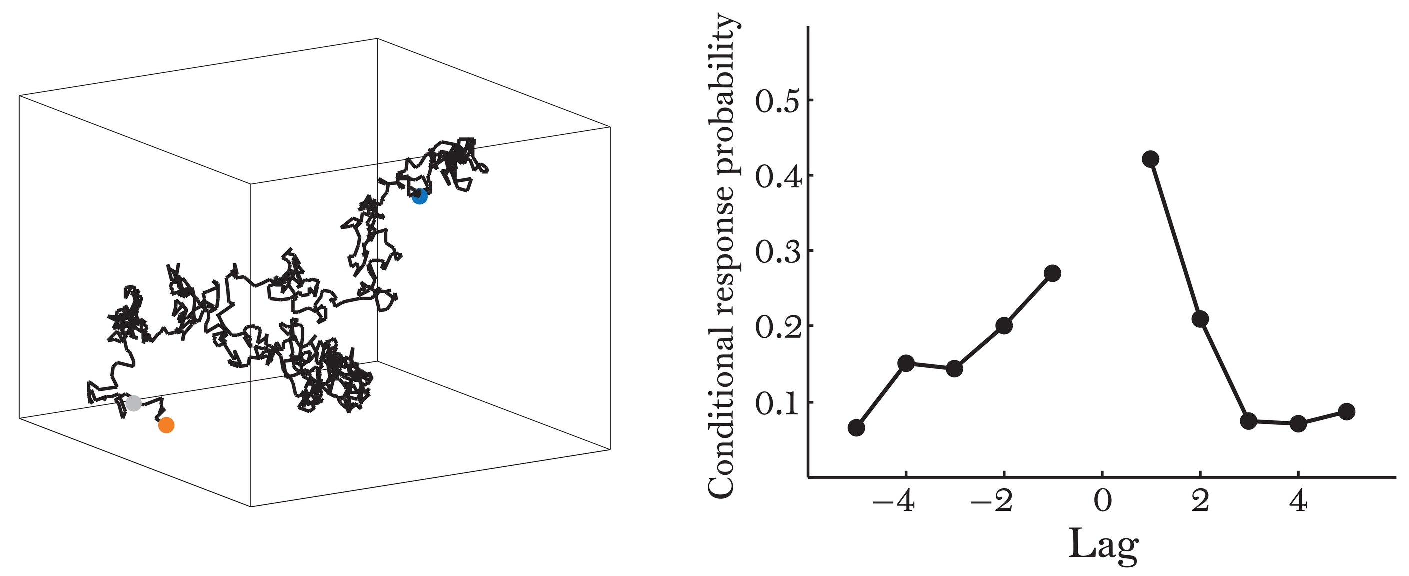
Jeremy R. Manning, David M. Blei, and Kenneth A. Norman
Princeton University

This research was supported by the NSF/NIH Collaborative Research in Computational Neuroscience Program, grant number NSF IIS-1009542.

Introduction

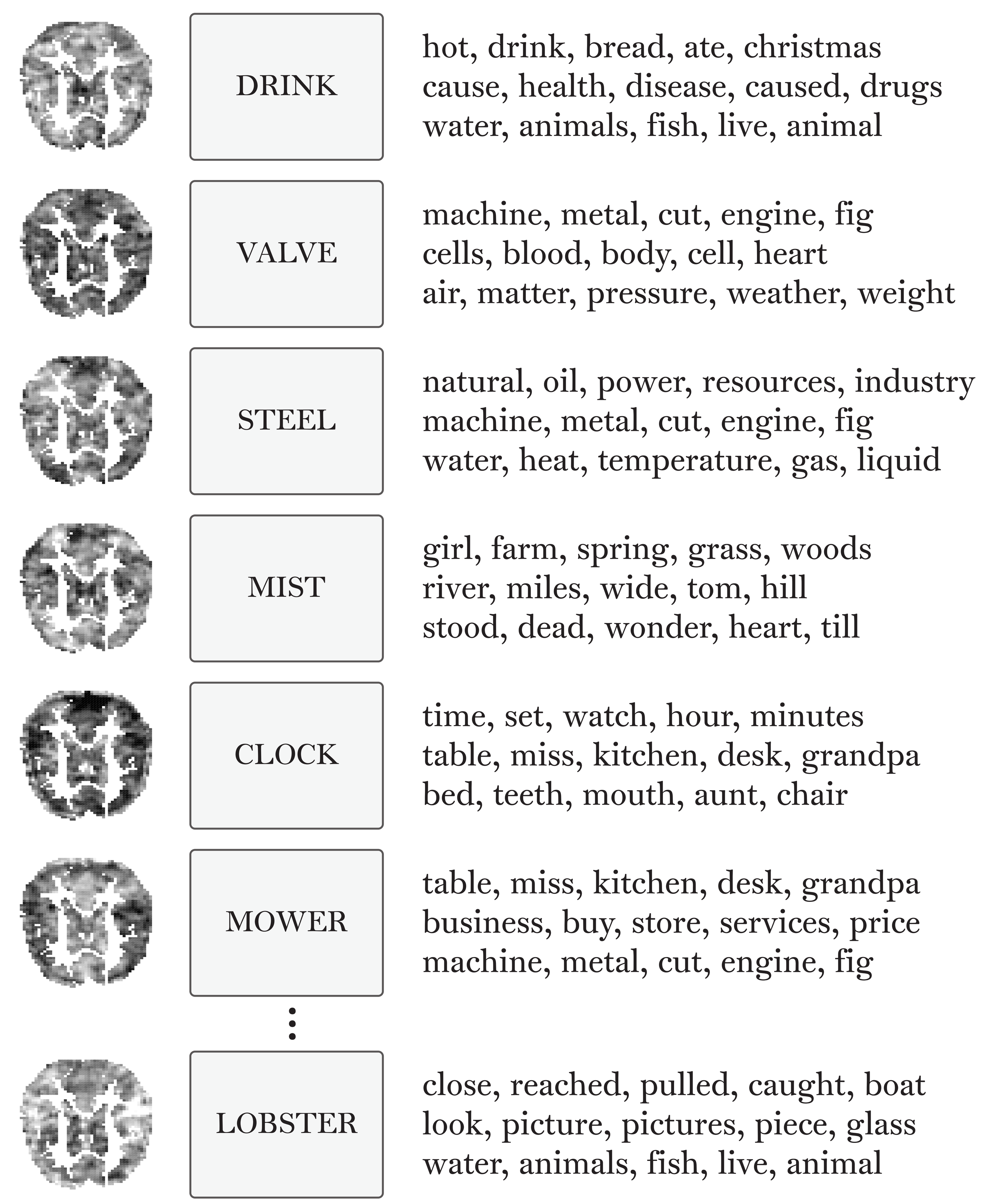
Context-based theories of memory posit that items on a studied list become associated with the mental context in which they are experienced.

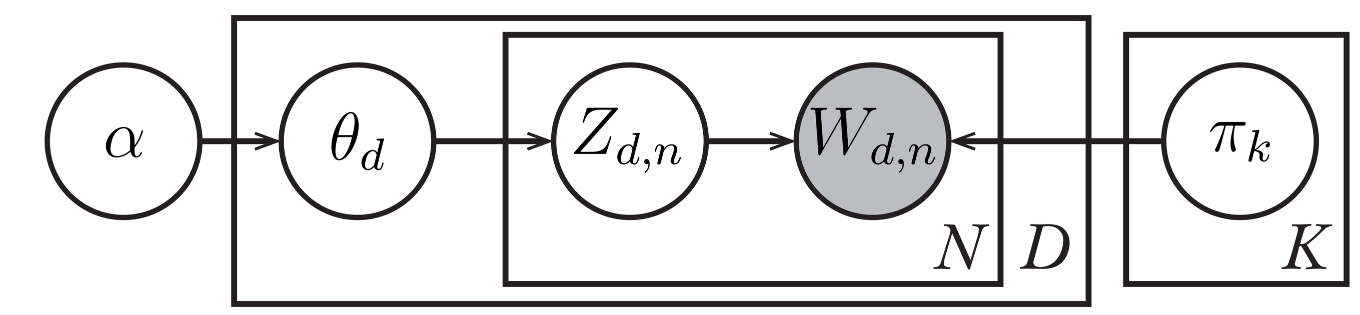
We present a framework for tracking the neural correlates of individual items and the contexts in which they are experienced, during individual presentation or recall events.



Context-based theories of memory. Context drifts gradually over time and becomes associated with each experienced event, giving rise to the contiguity effect in free recall. Our approach allows us to decode semantic information from neural patterns. This will allow us to examine the extent to which the evolving neural representation of context contains semantic information.

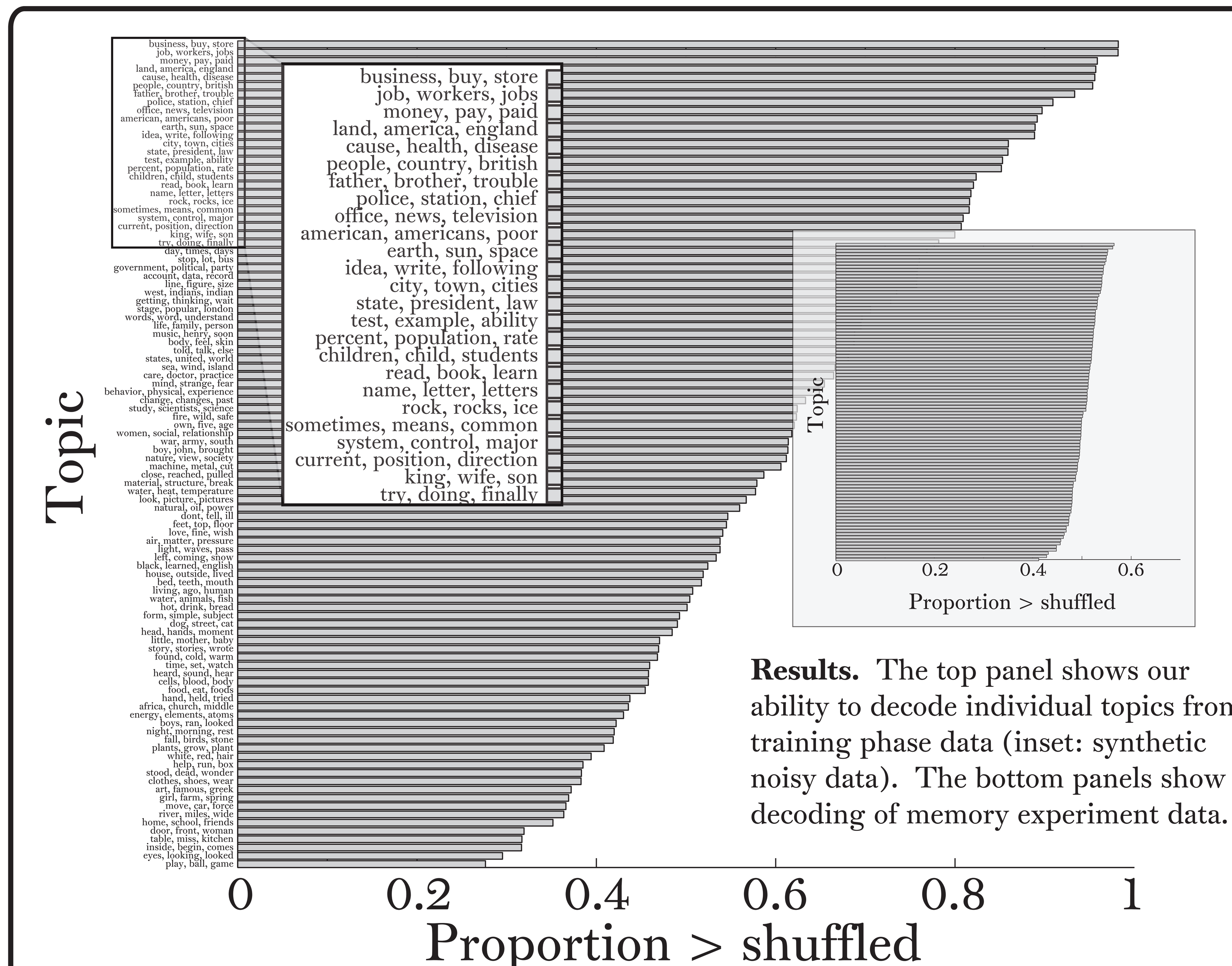
Overview. We measure the BOLD response evoked by each presented word during the experiment. We use Latent Dirichlet Allocation, a topic modeling algorithm, to compute topic vectors for each word in a large vocabulary. Using the known topic vectors for each of the presented words, our goal is to infer the neural representations of each topic. We then use the inferred topic representations to decode topic vectors from previously unseen neural patterns.



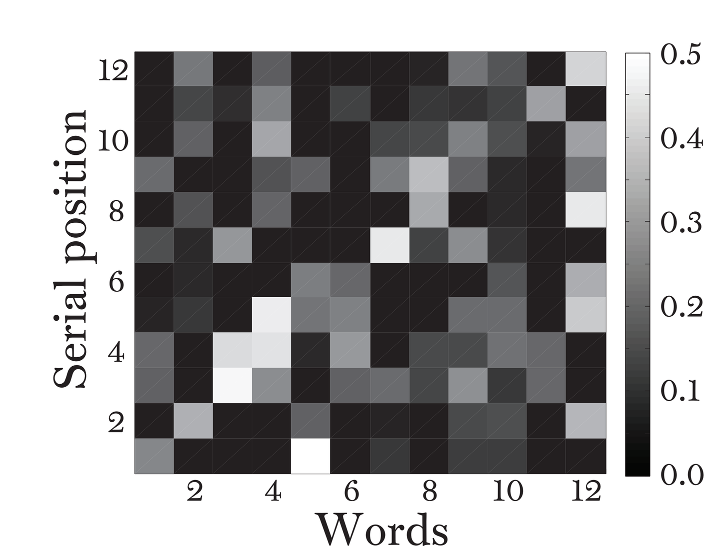


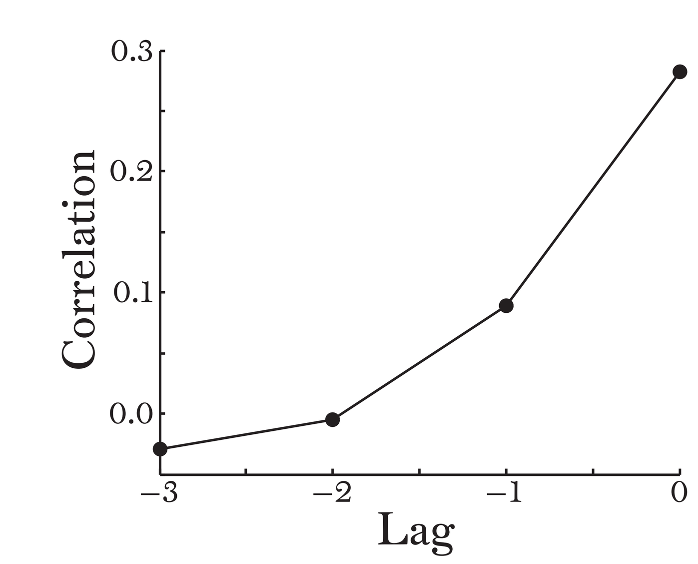
α	Topic sparsity parameter
θ_d	Topic proportions for document d
$Z_{d,n}$	Topic for word n in document d
$W_{d,n}$	Word n in document d
π_k	Topics (distributions over words)

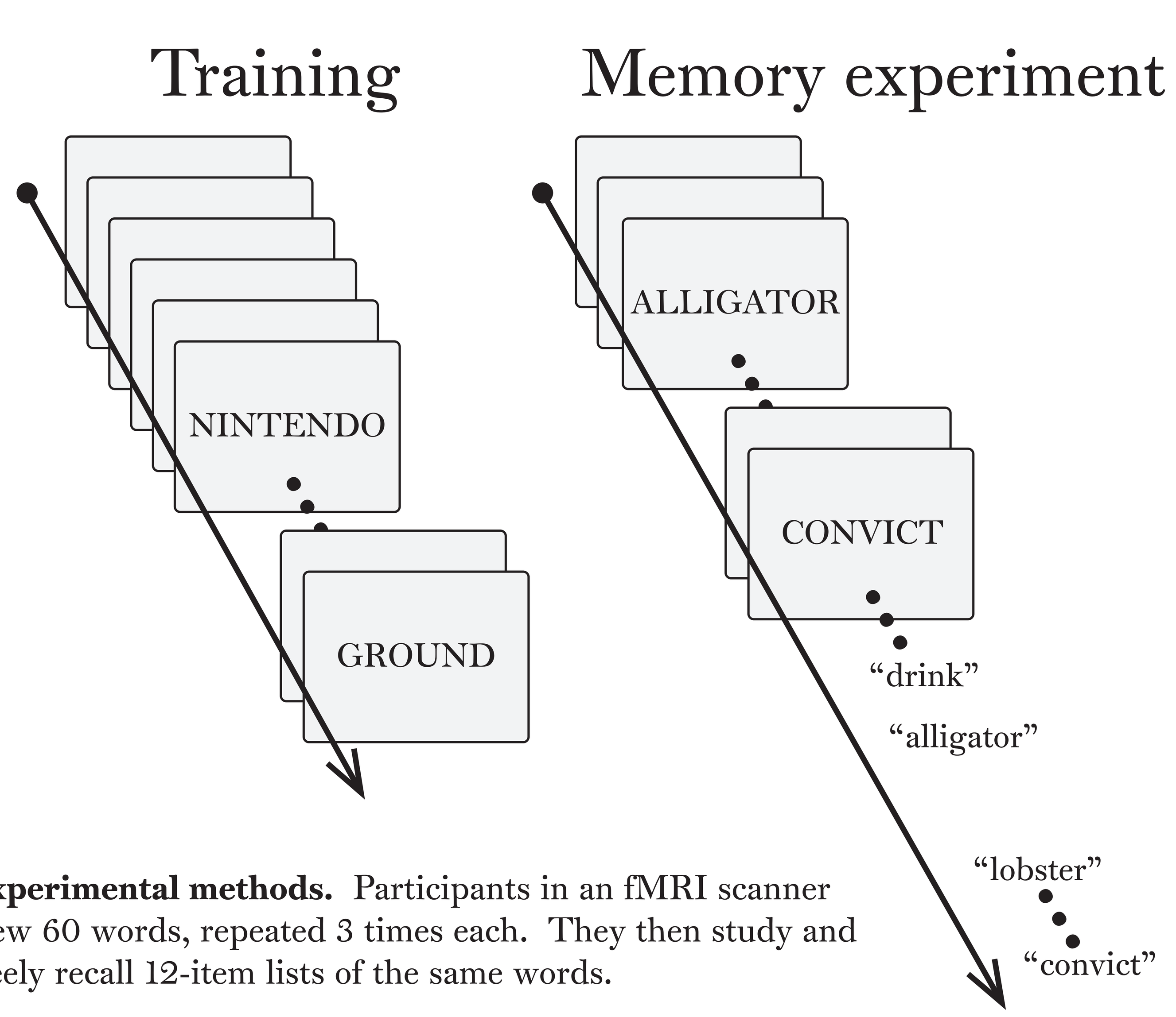
Latent Dirichlet Allocation. We assume each document is generated as follows, given a set of K topics: (1) select topics for the document according to the sparsity parameter; (2) for each word in the document, select the word's topic according to the topics for the document; (3) draw a word from that topic. LDA entails fitting the latent (unshaded) parameters given the observed (shaded) words in the documents.



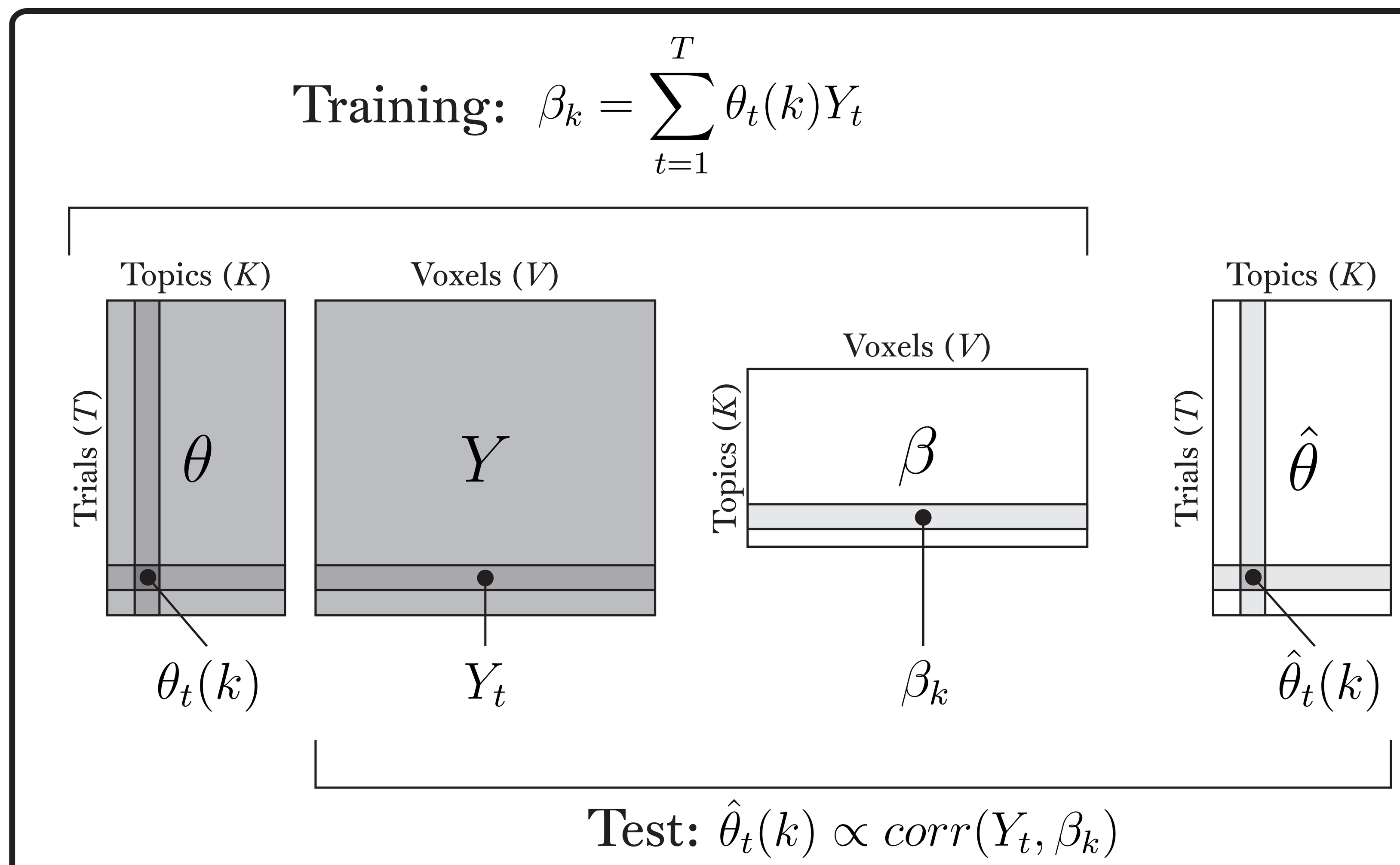
Results. The top panel shows our ability to decode individual topics from training phase data (inset: synthetic noisy data). The bottom panels show decoding of memory experiment data.







Experimental methods. Participants in an fMRI scanner view 60 words, repeated 3 times each. They then study and freely recall 12-item lists of the same words.



Decoding topic vectors from neural patterns. We use 5-fold cross validation to train and test our decoding algorithm. We infer the neural representations of each topic by computing weighted averages of the activity evoked by held-out words. We use correlations to decode topic vectors from in-fold neural patterns.