Factor Topographic Latent Source Analaysis: Factor Analysis for Brain Images

Jeremy R. Manning, Samuel J. Gershman, Kenneth A. Norman, and David M. Blei

Princeton University

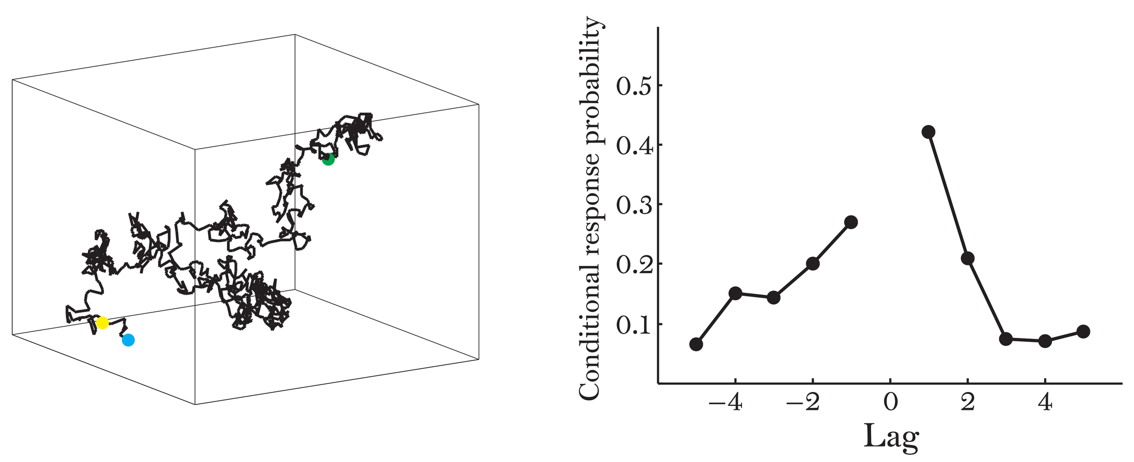
For reprints contact manning3@princeton.edu

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

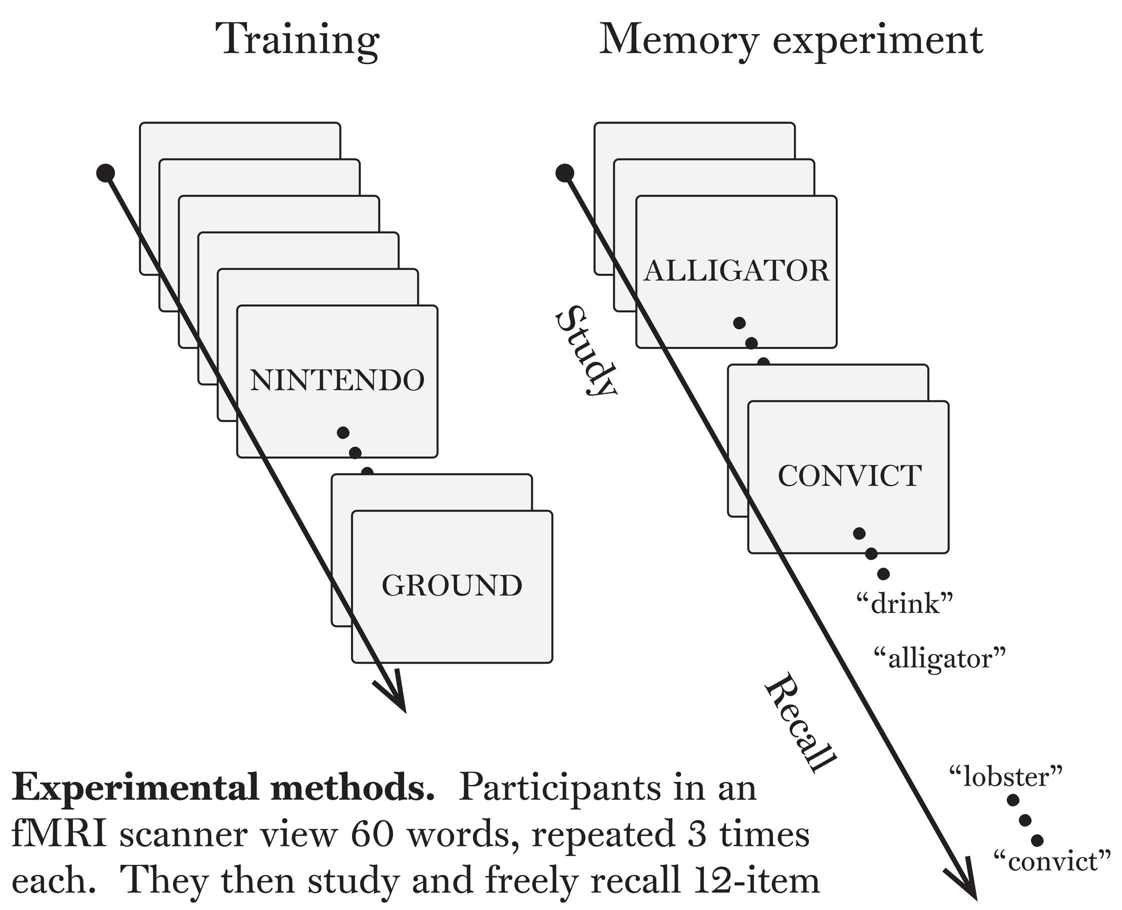
Overview & experimental methods

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

We present a framework for tracking the neural correlates of individual items[6] and the contexts in which they are experienced,[5] during individual study and recall events.



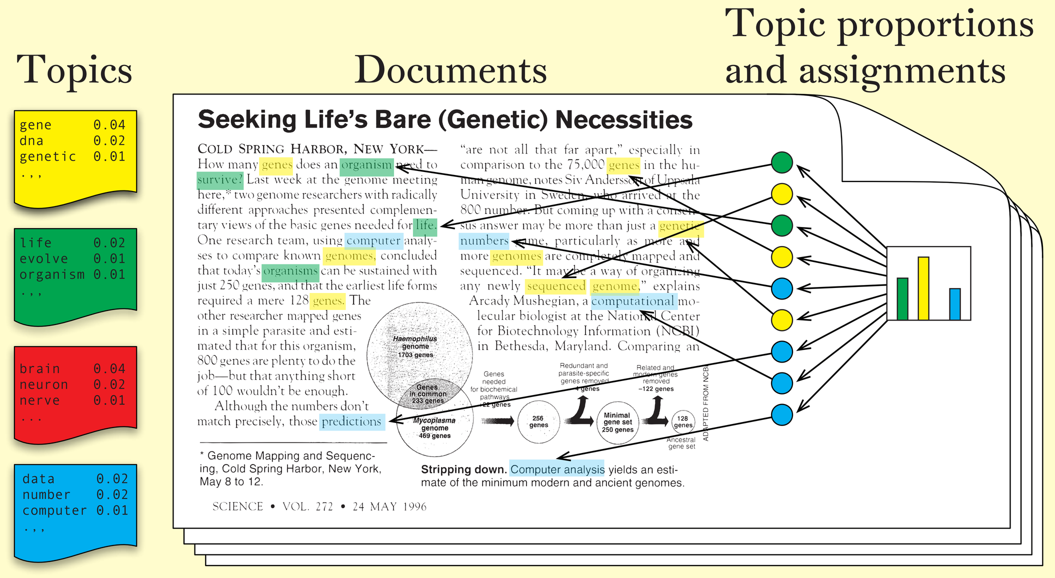
Context-based theories of memory. Context drifts gradually over time and becomes associated with each experienced event, giving rise to the contiguity effect in free recall.

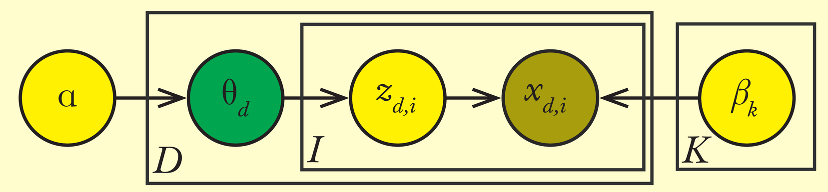


Experimental methods. Participants in an fMRI scanner view 60 words, repeated 3 times each. They then study and freely recall 12-item lists of the same words.

Towards a unified model of corpora and cognition. Our approach attempts to infer the evolving state of mental context using text, behavioral, and neural data.

Model of word meanings

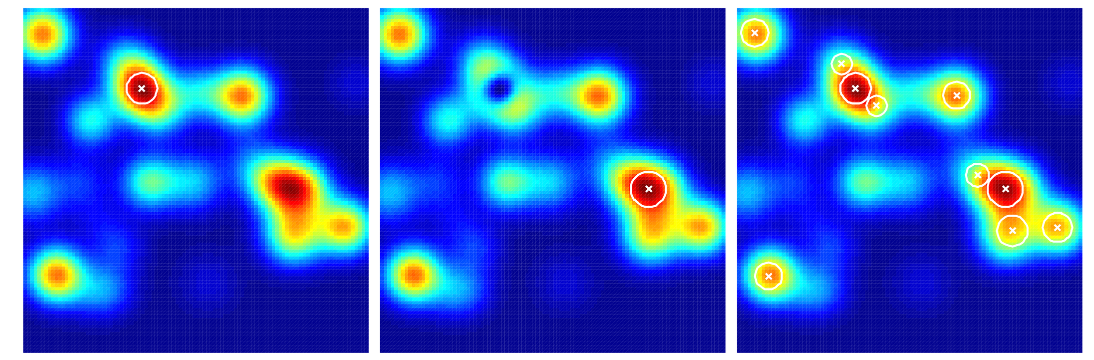




- α topic sparsity parameter
- θ_d topic proportions for document d
- $z_{d,i}$ topic for word i in document d
- $x_{d,i}$ word i in document d
- β_k topic k (a distribution over words)
- D number of documents
- I number of words in each document
- K number of topics

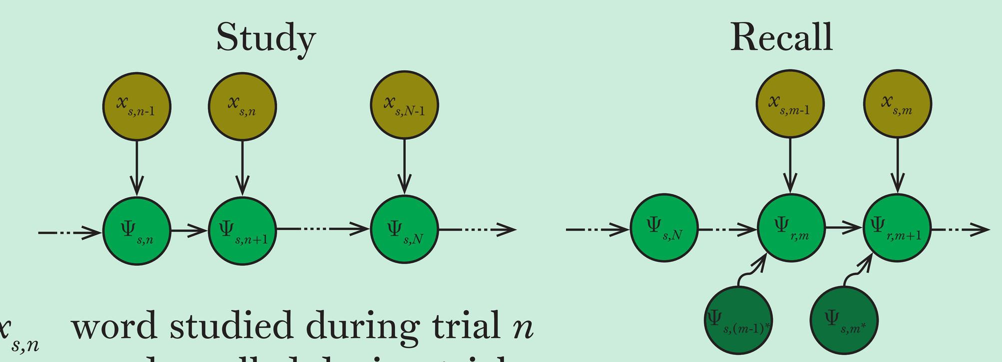
Latent Dirichlet Allocation. Topic vectors are derived by analyzing a collection of documents.[1] LDA entails fitting the latent (lightly shaded) variables given the observed (darkly shaded) words in the documents.

Fitting the models



Initializing source locations and widths. We use "hotspots" in the neural images to estimate the locations and widths of the sources. We can then solve for the source weights using linear regression. **Honing the parameter estimates.** We use a scalable stochastic variational inference-based fitting procedure to hone the parameter values given the observed neural and behavioral data.

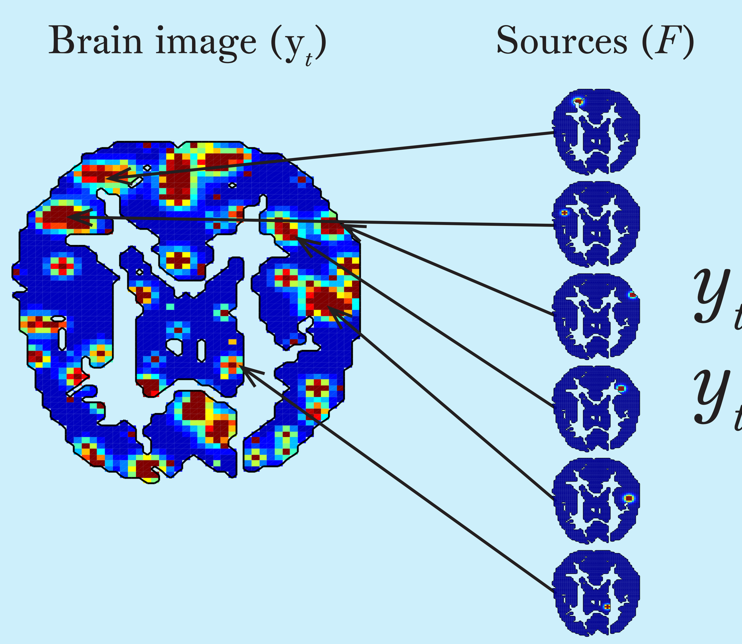
Model of mental context



- $x_{s,n}$ word studied during trial n
- $x_{r,m}$ word recalled during trial m
- $\Psi_{s,n}$ mental context during study of word n
- $\Psi_{r,m}$ mental context during recall of word m
- $\Psi_{s,m'}$ mental context associated with the m^{th} recalled word during study

Temporal Context Model. Mental context vectors are weighted averages of topic vectors for studied and recalled words.[3,8]

Model of neural activity

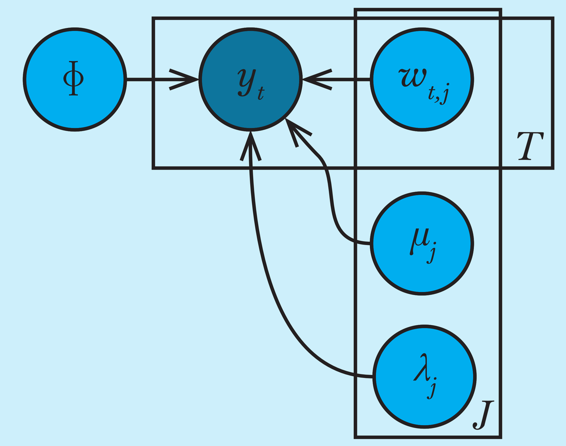


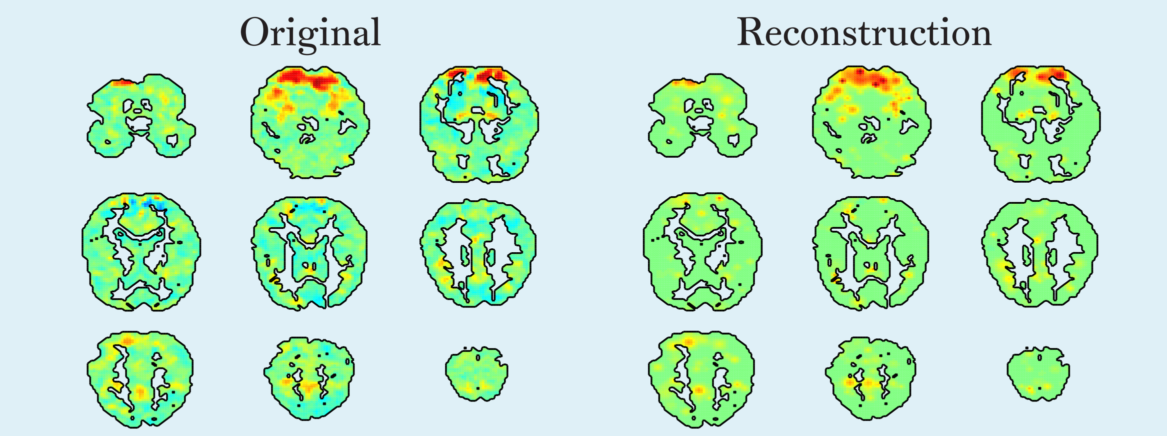
$$y_t = w_t F$$

$$y_t = \theta_t(w)^{\mathrm{T}} F$$

Topographic Latent Source Analysis. Neural patterns are represented as linear combinations of spherical sources.[2,4] Unsupervised (top) and supervised (bottom) equations are shown.

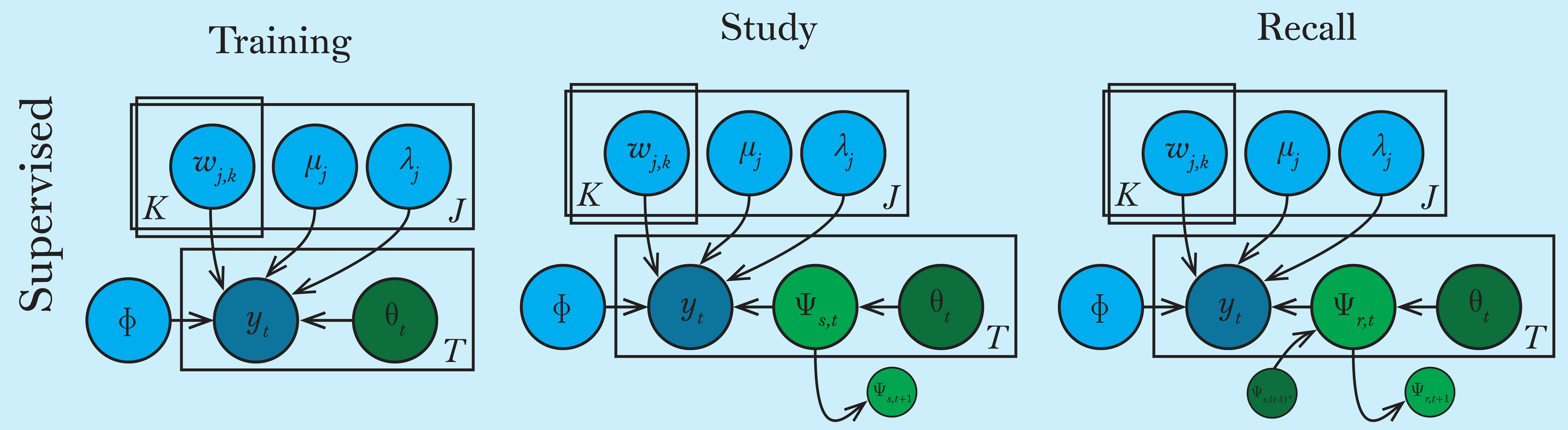
Unsupervised





Unsupervised reconstructions. Just a few hundred latent sources capture most of the variability in a 50,000-voxel image.

Supervised

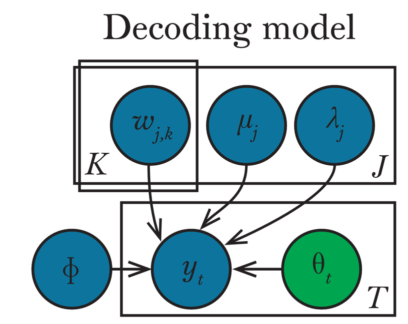


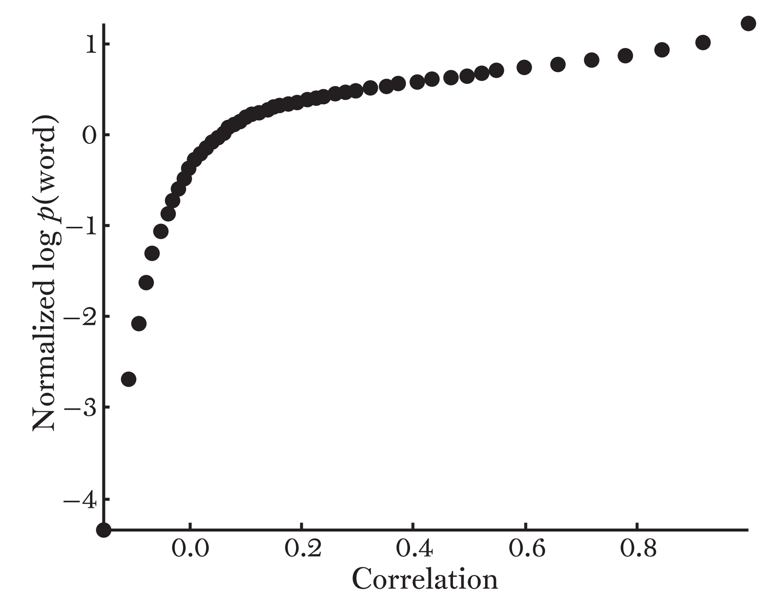
- ϕ voxel noise parameter
- y_t brain image during trial t
- $w_{t,j}$ source j's weight during trial t (unsupervised)
- $w_{j,k}$ source j's weight on topic k (supervised)
- θ_t topic vector for word experienced during trial t
- $\psi_{s,t}$ mental context during study of word t
- $\psi_{r,t}$ mental context during recall of word t
- μ_j source j's center location
- λ_j source j's width
- T number of trials
- J number of sources
- K number of topics

Models of evolving neural activity. The sources' center and width parameters are held fixed across trials. In the unsupervised model, the source weights vary independently to explain each image. In the supervised model, the source weights vary as a function of the semantic attributes of the word experienced as the images were collected. The study and recall models account for drifting mental context by introducing an additional autocorrelated latent variable. In the recall model, the mental context associated with the word being recalled is reinstated and influences the way in which mental context drifts.

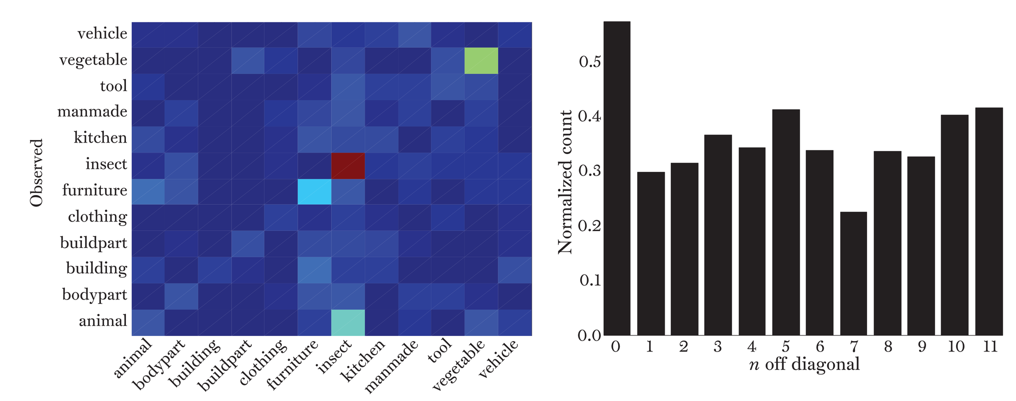
Results

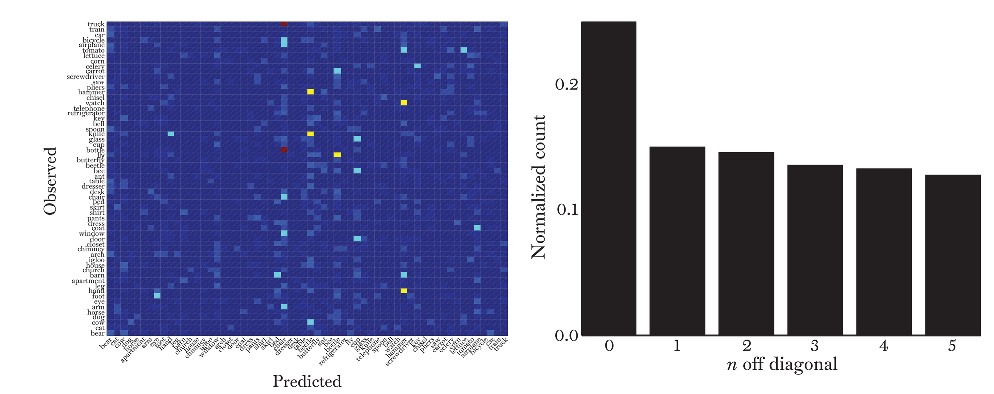
Decoding topic vectors from brain images. After fitting the model to training data[7], we treat the model parameters as observed and compute a posterior distribution over topic vectors (or mental context).





Posterior probability as a function of correlation. Correlations are between the actual word's topic vector and the topic vector for each alternative word.





Confusion matrices. These panels summarize decoding errors by category (top) and word (bottom). The right panels display the mean (normalized) counts as a function of the positional distance from the diagonals of the confusion matrices.

Bibliography

[1] Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. Journal of Machine Learning Research, 3: 993 - 1022.
[2] Gershman SJ, Blei DM, Pereira F, Norman KA (2011) A topographic latent source model for fMRI data. NeuroImage, 57: 89 - 100.
[3] Howard MW, Kahana MJ (2002) A distributed representation of temporal context. Journal of Mathematical Psychology, 46: 269 - 299.
[4] Manning JR, Norman KA, Blei DM (2012) Factor topographic latent source analysis: Factor analysis for brain images. Submitted.
[5] Manning JR, Polyn SM, Baltuch G, Litt B, Kahana MJ (2011) Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. Proceedings of the National Academy of Sciences, USA, 108: 12893 - 12897.
[6] Manning JR, Sperling MR, Sharan A, Rosenberg EA, Kahana MJ (2012) Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. The Journal of Neuroscience, 32: 8871 - 8878.
[7] Mitchell TM, Shinkareva SV, Carlson A, Chang K-M, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. Science, 320: 1191 - 1195.
[8] Socher R, Gershman SJ, Perotte AJ, Sederberg PB, Blei DM, Norman KA (2009) A Bayesian analysis of dynamics in free recall. Advances in Neural Information Processing Systems, 22: 1714 - 1722.