How does the brain represent and retrieve word meanings?

Jeremy R. Manning & Michael J. Kahana

Princeton University, University of Pennsylvania

Overview

We studied how the brain represents, organizes, and retrieves conceptual (semantic) information.

We examined recordings from neurosurgical patients as they studied and freely recalled lists of words.

Whereas previous studies have attempted to decode stimulus identities from neural patterns, our goal was to infer how stored representations of stimuli are organized in participants' memories.

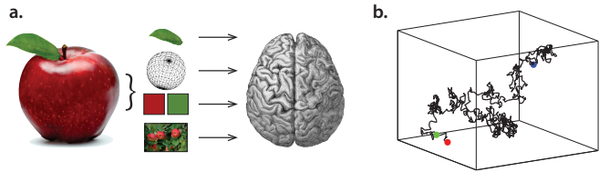

Figure 1. *Representing and retrieving conceptual memories.* **a.** *Distributed memory hypothesis.* Concepts are broken down into constituent semantic features; each feature is represented by a distributed pattern of brain activity. **b.** *Contextual reinstatement hypothesis.* Context drifts gradually over time and is associated with each experienced event.

Methods

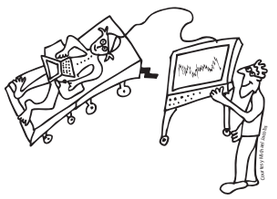

Figure 2. *Our recording setup.* Patients are implanted with subdural and depth electrodes. Experiments are administered on a bedside laptop computer.

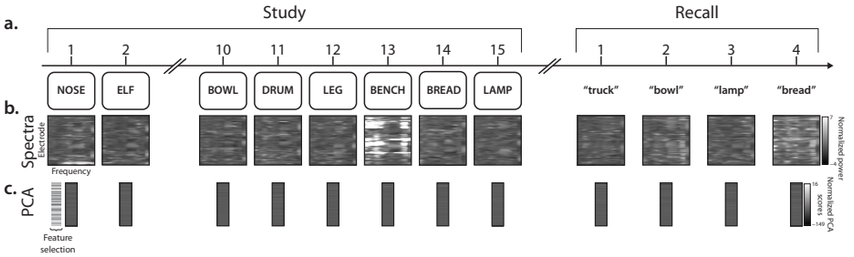

Figure 3. *Analysis.* **a.** Participants study and freely recalls lists of 15 or 20 common nouns. **b.** For each electrode we compute mean power contained in 50 frequencies (2 - 99 Hz) during each study and recall event. **c.** We apply PCA and select features that either vary with the meanings of studied words or vary gradually during the study interval. We examine the selected features during the recall interval.

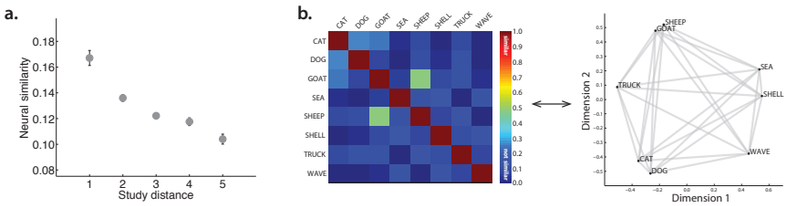

Figure 4. *Feature selection.* **a.** *Contextual feature selection.* We compute the neural similarity between each pair of studied words, for each neural component. Correlations between neural and temporal similarity imply that the given neural component evolves gradually during study. **b.** *Semantic feature selection.* We compute the neural and semantic similarity between each pair of studied words. (A seperate neural similarity matrix must be constructed for each neural component.) Correlations between neural and semantic similarity imply that the given neural component represents the meanings of the studied words.

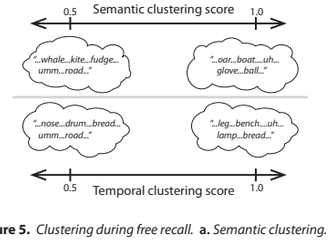

Figure 5. *Clustering during free recall.* **a.** *Semantic clustering.* The semantic clustering score measures a participant's tendancy to successively recall semantically related words. **b.** *Temporal clustering.* The temporal clustering score measures a participant's tendancy to successively recall words that appeared at nearby positions on the studied list.

Results

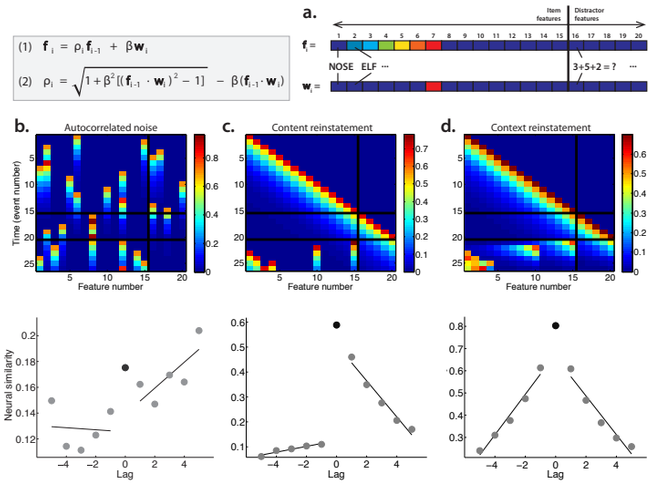

Figure 6. *Predicted neural similarity as a function of lag according to three models.* **a.** In each simulation, a single neuron is activated during each experimental event, *i*. Once activated, a neuron's activity decays gradually according to equations (1) and (2). **b.** *Autocorrelated noise.* Each experimental event activates a random neuron, irrespective of which item is being presented or recalled. **c,d.** Each neuron is activated by a single item or distractor during study. **c.** *Content reinstatement.* During recall of the j^{th} presented item we set $\mathbf{f}_i = \mathbf{w}_j$. **d.** *Context reinstatement.* During recall of the j^{th} presented item we set $\mathbf{f}_i = \mathbf{f}_j$.

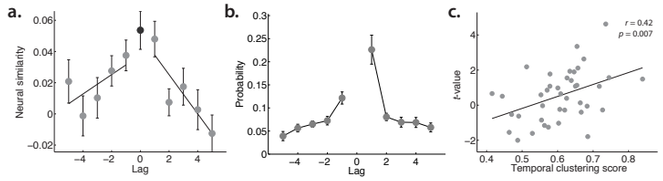

Figure 7. *Neural signature of contextual reinstatement.* **a.** Observed neural similarity between the feature vector corresponding to recall of a word from serial position *i* and study of a word from serial position *i*+lag. **b.** Probability of recalling an item from serial position *i*+lag immediately following an item from serial postion *i*, conditional on the availability of an item in that list position for recall. **c.** Participants exhibiting greater context reinstatement also exhibited more pronounced contiguity effects.

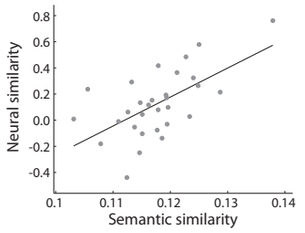

Figure 8. *Neural vs. semantic similarity.* Neural activity is recorded just prior to each recall. The dots indicate the means of 30 equally sized bins.

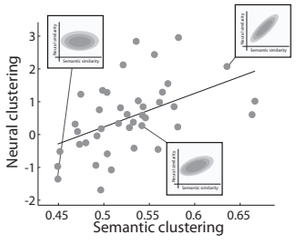

Figure 9. *Neural vs. semantic clustering.* Each dot represents a single participant. Neural clustering indicates the *t*-value from the correlation between neural and semantic similarity. (Data are from PFC.)

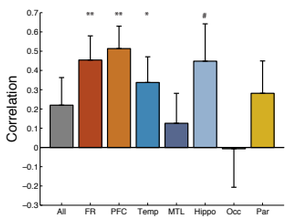

Figure 10. *Regions that predict semantic clustering.* Each bar indicates the correlation between neural and semantic clustering from the indicated region of interest.

Conclusions

We identified components of brain activity that varied with the meanings of words during study. We examined these same "semantic features" just prior to recall.

The correlation between neural and semantic similarity was preserved in prefrontal, occipital, and lateral temporal cortex, indicating that these regions are involved with representing the meanings of words.

We also found that the correlation between neural and semantic similarity predicted semantic clustering. The prefrontal cortex, lateral temporal cortex, and hippocampus exhibited this effect, indicating that these regions are involved with organizing the memories of the words.

We also identified components of brain activity that evolved gradually during study. These "contextual features" were reinstated just prior to recall.

The magnitude of the neural signature of contextual reinstatement predicted participants' tendencies to temporally cluster their recalls.

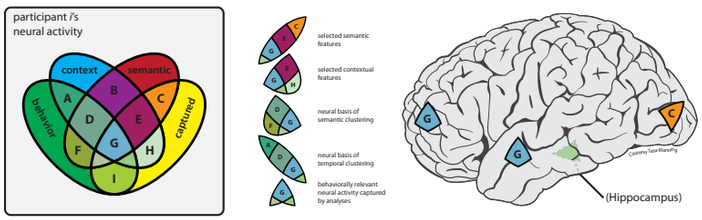

Figure 11. *Interpreting components of neural activity.* The box represents the full pattern of brain activity exhibited by participant *i* during the experiment. Our feature selection framework identifies neural patterns in segments C, E, G, and H. Neural patterns in segment A, D, F, and G contain information about the order in which participants will recall the words.

For reprints contact **manning3@princeton.edu**