Pattern classification of fMRI retrieval states in recognition memory

Joel R. Quamme, David J. Weiss, & Kenneth A. Norman

Department of Psychology, Princeton University, NJ 08540

Princeton Computational Memory Lab

2070

Introduction

There is general agreement that recognition memory can be based on (1) assessments of stimulus familiarity, or (2) recollection of specific details about a prior event.

However, there is still significant debate over when and how the two are used to make recognition decisions.

We used *pattern classification of distributed functional MRI activity* (e.g., Polyn et al., 2005) to extract neural information about people's use of recollection or familiarity, and we relate this information to recognition behavior.

KEY THEORETICAL ISSUE: Attention to recollection and familiarity may be subserved by regions in the parietal lobe (Wagner et al., 2005). However, it is unclear whether existing evidence favors (1) a state of trying to recollect, or (2) evaluating retrieved information.

We asked:

1. Can we find evidence of *differential strategic modes* in the brain associated with trying to recollect or trying to judge familiarity?
2. Does neural information about strategic mode allow us to predict recognition behavior in theoretically-interesting ways?
3. Can we test theories about the brain systems involved in memory retrieval?

Two-Phase Design

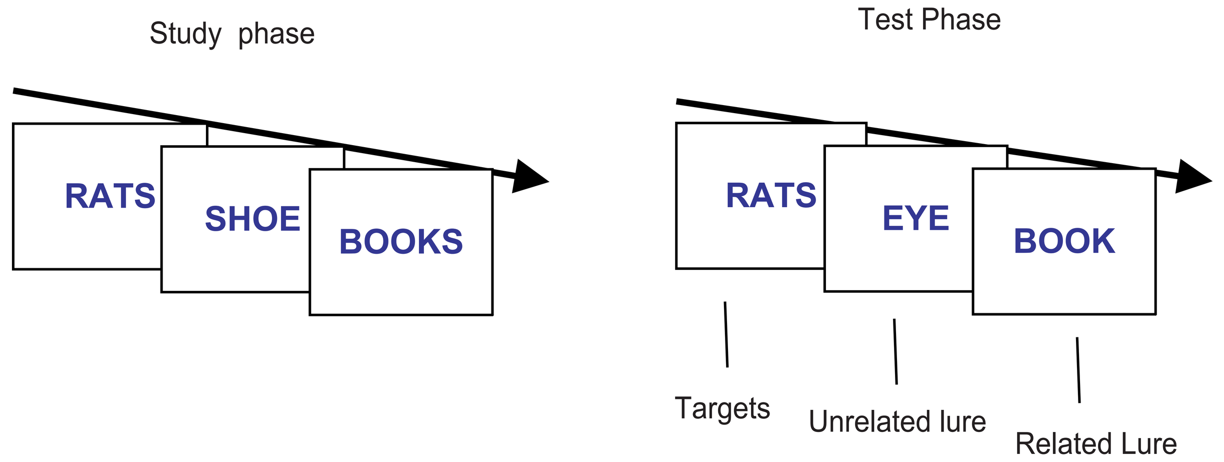
PHASE 1:

- (i) Experimentally elicit two strategic retrieval states
- (ii) Train a classifier to distinguish the states based on brain activity

PHASE 2:

- (iii) Apply the trained classifier to brain activity on a different task
- (iv) Read out, trial-by-trial, the classifier's estimate of the subject's cognitive state.
- (v) Relate classifier output to behavior.

Plurals Paradigm (Hintzman, Curran, & Oppy, 1992)



Switched-plural lures are related to study items, so should be familiar; but, they can be rejected if the study item is recollected.

Can we predict related lure responses from brain activity using a classifier that has been trained to distinguish "recollection mode" from "familiarity mode"?

Methods

12 Subjects
3-T Siemens Allegra magnet; 34 axial slices; TR=2000ms; voxel size: 3x3x3mm
1 hour training on all task procedures before the scan
1 hour session inside the scanner

Phase 1: Old/New recognition

- Study: 24 concrete singular and plural words in each of 5 study-test cycles (120 total)
- Test: 48 items per cycle (24 old, 24 new), 24 appearing in each of two blocked conditions:
  - Instruct subjects to judge whether they recollect something specific about the item.
  - Instruct subjects to judge quickly whether the item is familiar.

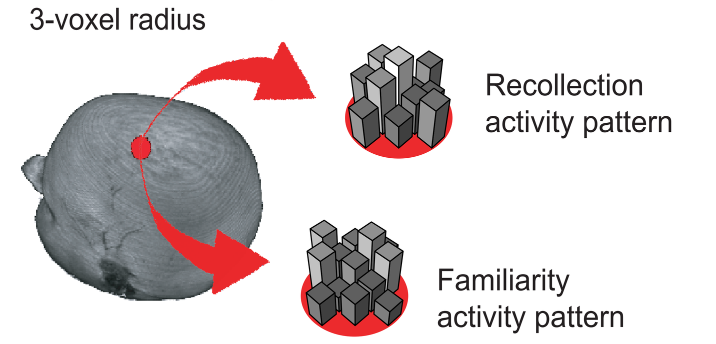
Logistic Regression Classifier

Feed in voxel activity.

Train the classifier to distinguish brain patterns associated with attempting to recollect vs. attempting to judge familiarity; classifier learns a mapping between brain activity patterns and conditions.

Regularization parameter to guard against overfitting -- penalizes regression solutions for weights that are too high.

Local Pattern Mapping (Kriegeskorte et al., 2006)



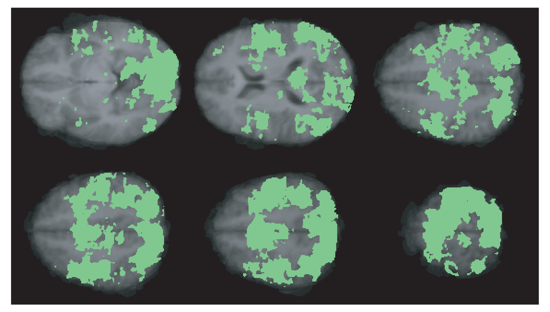
1. sweep the searchlight through the brain
2. apply classifier to pattern of voxel activity within each searchlight
3. create brain maps showing spheres that carry the most information about recollection

N-1 Cross-validation

We need to know where the classifier is learning something *reliable* about how the two conditions differ

Train on 4 of the 5 runs, test generalization to the remaining run

Widespread above-chance classifiability at $p < .005$ across subjects



Lots of areas carry information about about the difference between recollection and familiarity modes.

But which spheres are carrying behaviorally-relevant information?

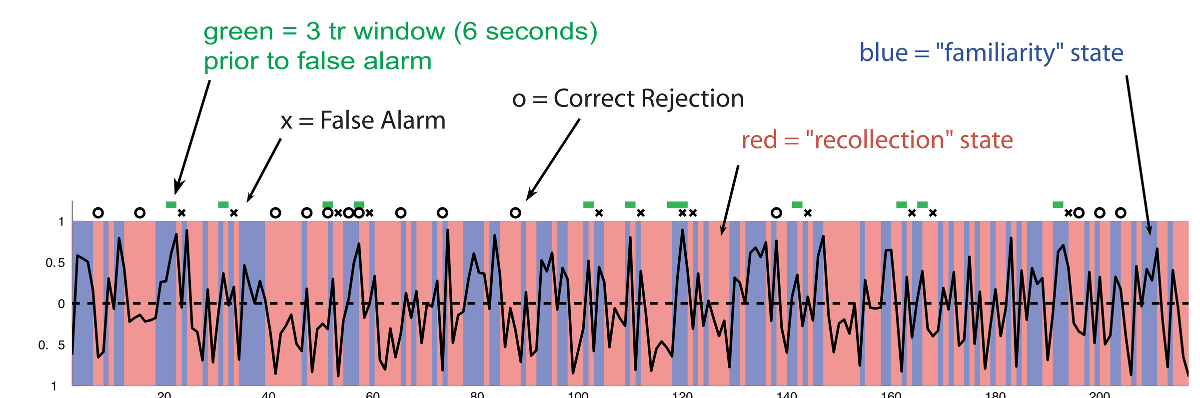
Phase 2: Old/Related/New recognition

Study 52 singular and plural words in each of 3 study-test cycles
Test 26 old, 26 new, and 26 switched-plural related lures in each cycle
Event-related design; jittered with 26 null trials (ISI= 4-12 seconds)

Does the classifier track behavior in a theoretically-meaningful way?

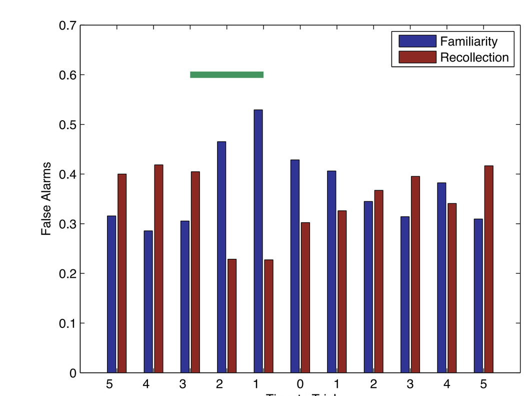
1. Train on all phase 1 data; apply trained classifier to spheres in phase 2 data (example sphere from 1 subject)

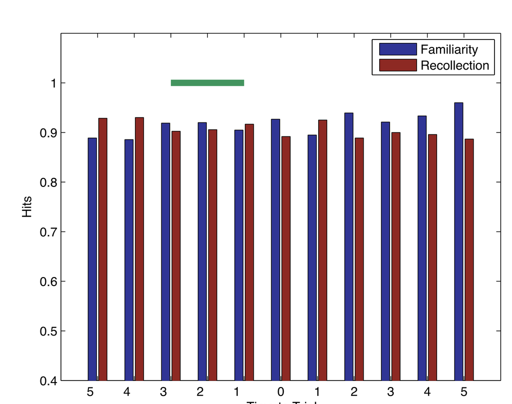
2. Compute time course of classifier's "best guess" as to whether activity at each time point more closely matches recollection or familiarity:



3. Bin subject's responses by classifiers output at time points before, during, and after stimulus and compute the strength of 2 theoretically meaningful response patterns

We have a clear expectation that recollection states should be associated with fewer related lure false alarms. Recollection state (vs. familiarity state) should not have a strong effect on hits.

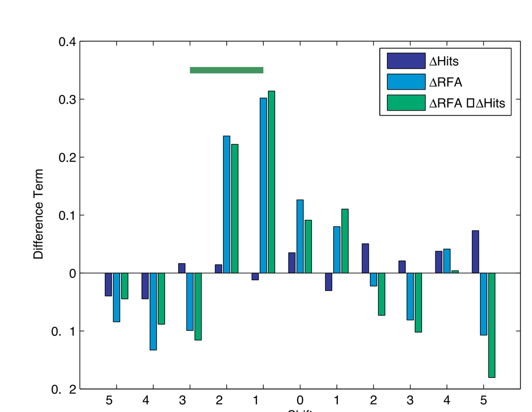




4. Find spheres show the desired effects for

(i) a 3 TR (6 second) window before the stimulus (as indicated by the green bar on figures)

(ii) a 3 TR window after the stimulus



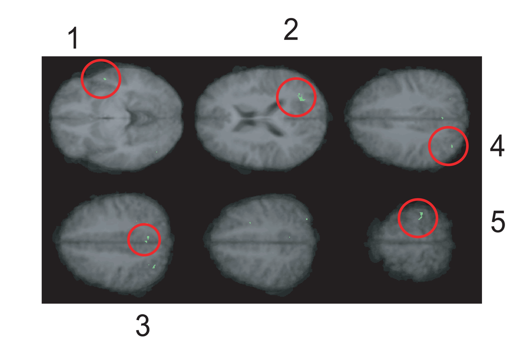
5. Select spheres for pre-trial and post-trial windows that meet $p < .005$ across subjects for:
- (i) cross-validation accuracy > 0.5
- (ii) recollection related lure FAs < familiarity related lure FAs
- (iii) one-way interaction of classifier state and trial such that FA difference > HR difference
  - (familiarity FAs - recollection FAs) - (familiarity HR - recollection HR)

This work was supported by NIMH grants R01 MH069456 and P50 MH062196

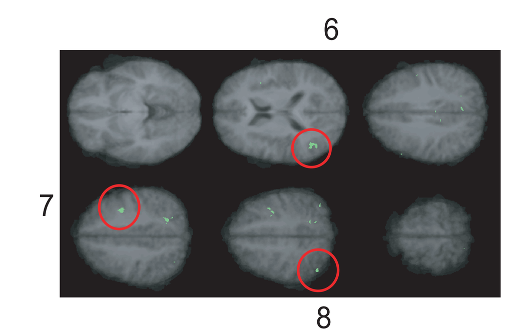
This poster can be downloaded from http://compmem.princeton.edu/publications.html

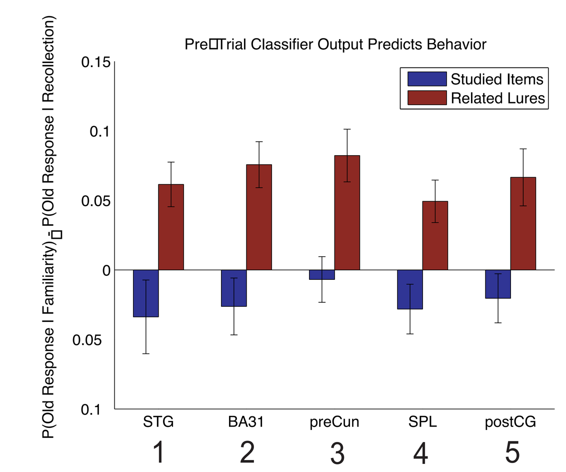
Areas predicting a specific decrease in related lure false alarms when the classifier identifies a recollection state as opposed to a familiarity state:

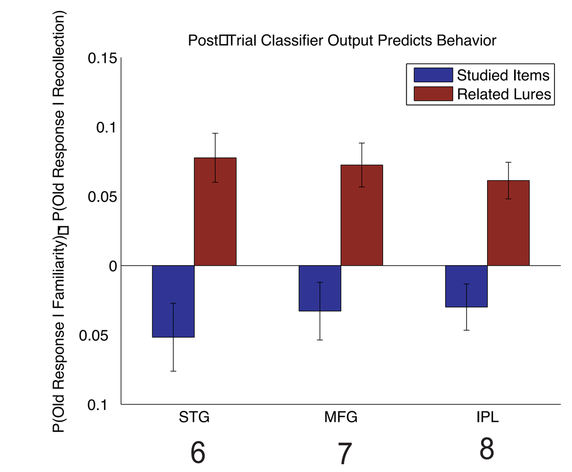
Before the item appears:



After the item has appeared:







Pre-trial prediction suggests the following areas are involved in trying to recollect (listening for recollection):
1. Superior Temporal Gyrus
2. Precuneus/Brodmann area 31
3. Precuneus
4. Superior parietal Lobule/Inferior Parietal Sulcus
5. Postcentral Gyrus

Post-trial prediction suggests the following areas are involved in evaluating recollected information:
6. Superior Temporal Gyrus
7. Middle Frontal Gyrus
8. Inferior Parietal Lobule/Inferior Parietal Sulcus

Discussion

Pattern classification can measure and distinguish states elicited by trying to recollect as opposed to relying exclusively on familiarity.

Classifier measures show fluctuations across time in "recollection mode" vs. "familiarity mode" during unconstrained testing that predict related-lure false alarms.

Different set of regions predicted related-lure false alarms before the stimulus and after the stimulus

There is a high potential for using neural classification measures to inform theories of memory.

References

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory: Registration without learning? Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 667-680.

Kreigeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. Procedings of the National Academy of Sciences of the United States of America, 103, 3863-3868.

Polyn S.M., Natu V.S., Cohen J.D., & Norman K.A. (2005) Category-specific cortical activity precedes recall during memory search. Science, 310, 1963-1966.

Wagner, A. D., Shannon, B. J., Kahn, I., Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. Trends in Cognitive Science, 9, 445-453