



lewispeacock@wisc.edu

# Probing Working Memory Representations with Pattern Classification



THE UNIVERSITY OF WISCONSIN MADISON

Jarrold A. Lewis-Peacock and Bradley R. Postle  
Department of Psychology, University of Wisconsin - Madison

## Introduction

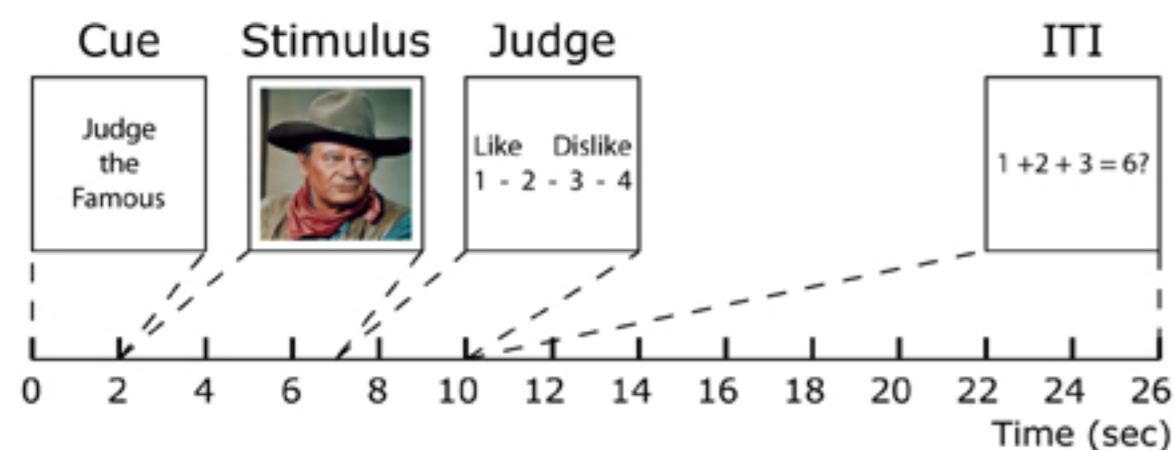
- This study effected a decisive test of the hypothesis that the **temporary activation of LTM representations** contributes to the retention of information in working memory. (e.g., Anderson, 1983; Cowan, 1995; Oberauer, 2002; Ruchkin et al., 2003).
- Several previous studies have produced data consistent with the **activated LTM model** (e.g., Druzgal and D'Esposito, 2003; Postle et al., 2003; Ranganath et al., 2004; Postle, 2005), but such results cannot be interpreted as direct tests of this model because of the problem of 'reverse inference' (Poldrack, 2006).
- This experiment was designed to support stronger inference than these earlier studies by restricting the ability to interpret delay-period activity to the recognition of specific patterns of brain activity from within specific brain regions that were previously identified during engagement of LTM processes.

We addressed 4 questions in this study:

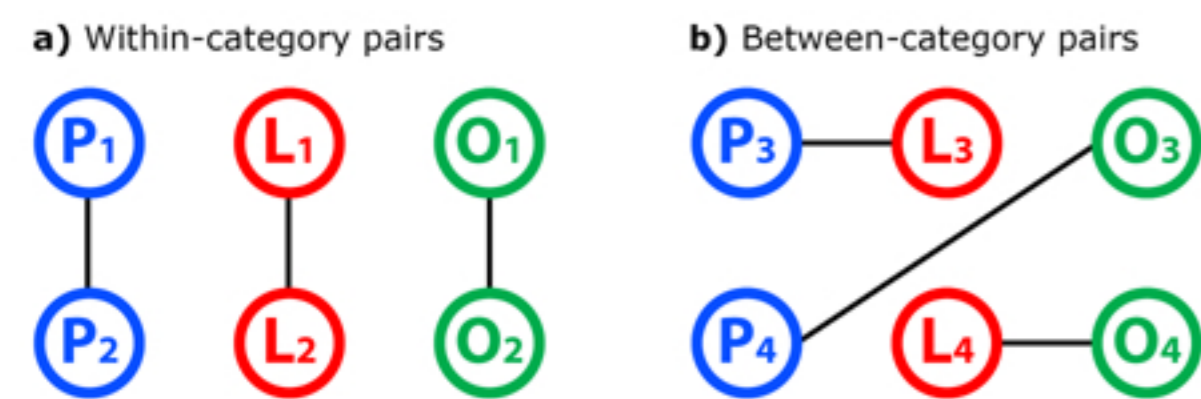
- Can delay-period activity be decoded with a LTM-trained pattern classifier?  
(... if so, then ...)
- What is the anatomical distribution of voxels supporting classification?
- Are neural representations *localized* or *distributed*?
- Are neural representations *linear* or *conjunctive*?

## Experiment

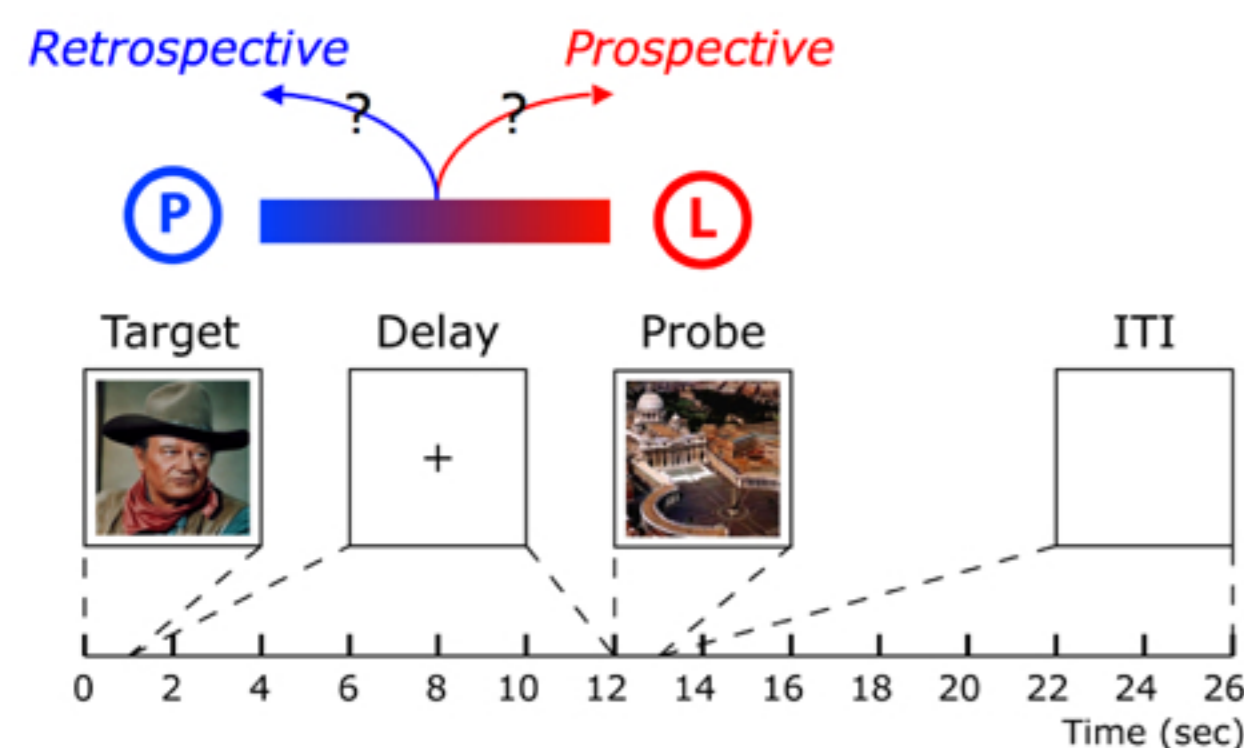
1. Stimulus-judgment task (engages LTM; Polyn et al. 2005)



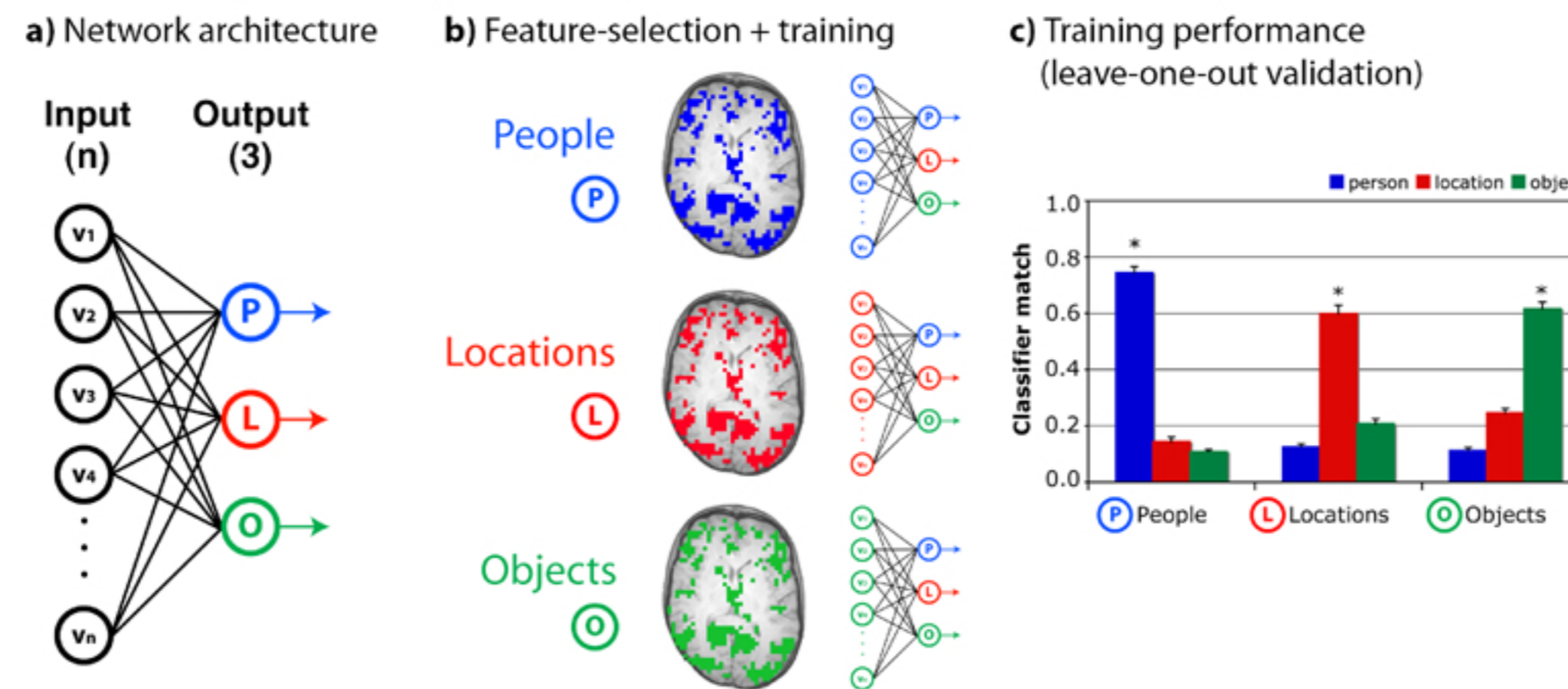
2. Learn arbitrary stimulus pairings of **People, Locations, & Objects**



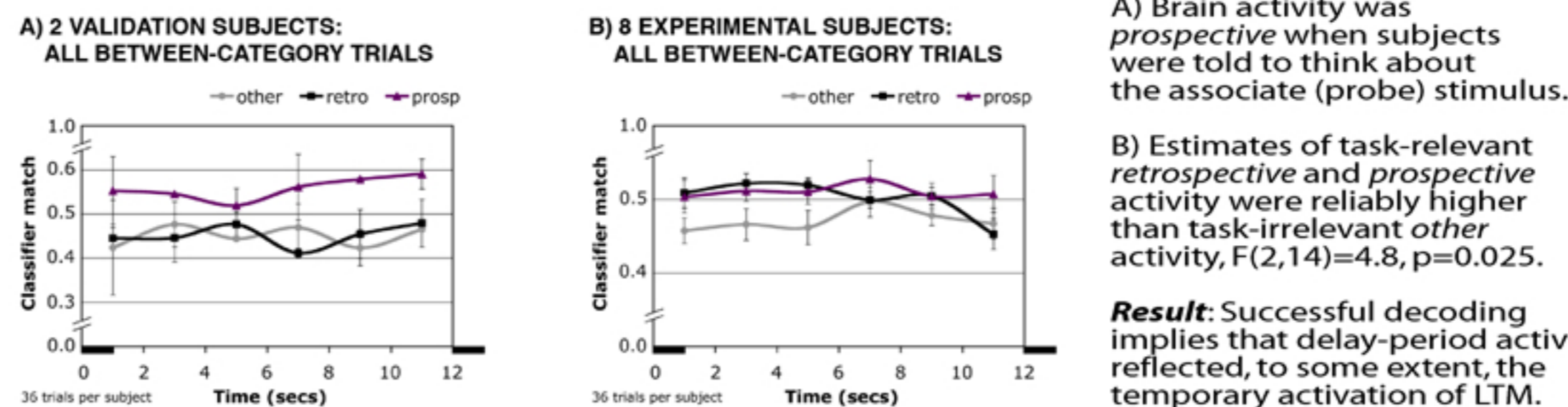
3. Delayed paired-associate recognition (requires WM retention)



## Training a classifier with LTM activity



### 1 Can we decode delay-period activity?



A) Brain activity was *prospective* when subjects were told to think about the associate (probe) stimulus.

B) Estimates of task-relevant *retrospective* and *prospective* activity were reliably higher than task-irrelevant *other* activity,  $F(2,14)=4.8, p=0.025$ .

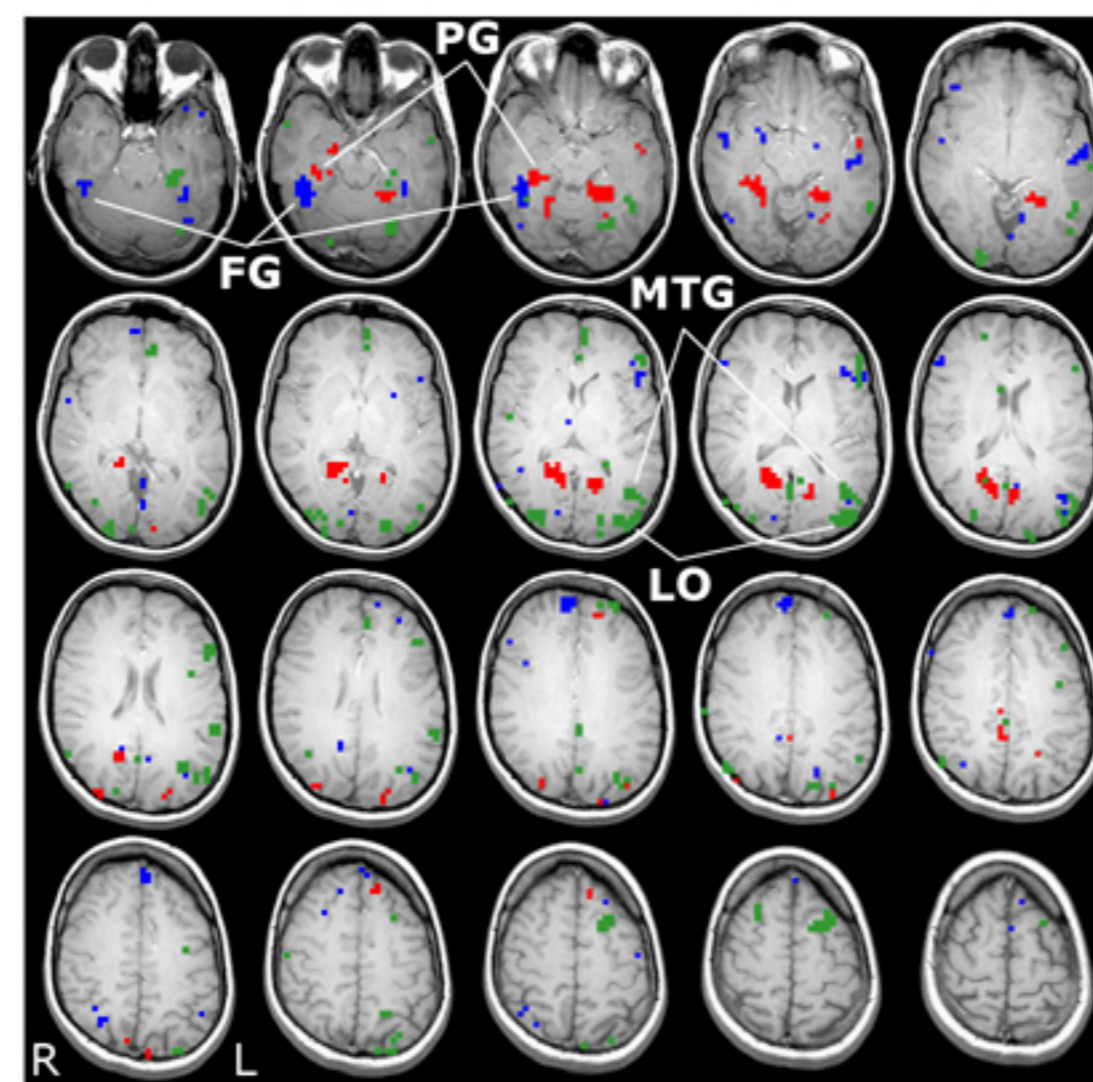
**Result:** Successful decoding implies that delay-period activity reflected, to some extent, the temporary activation of LTM.

### 2 Which voxels support classification?

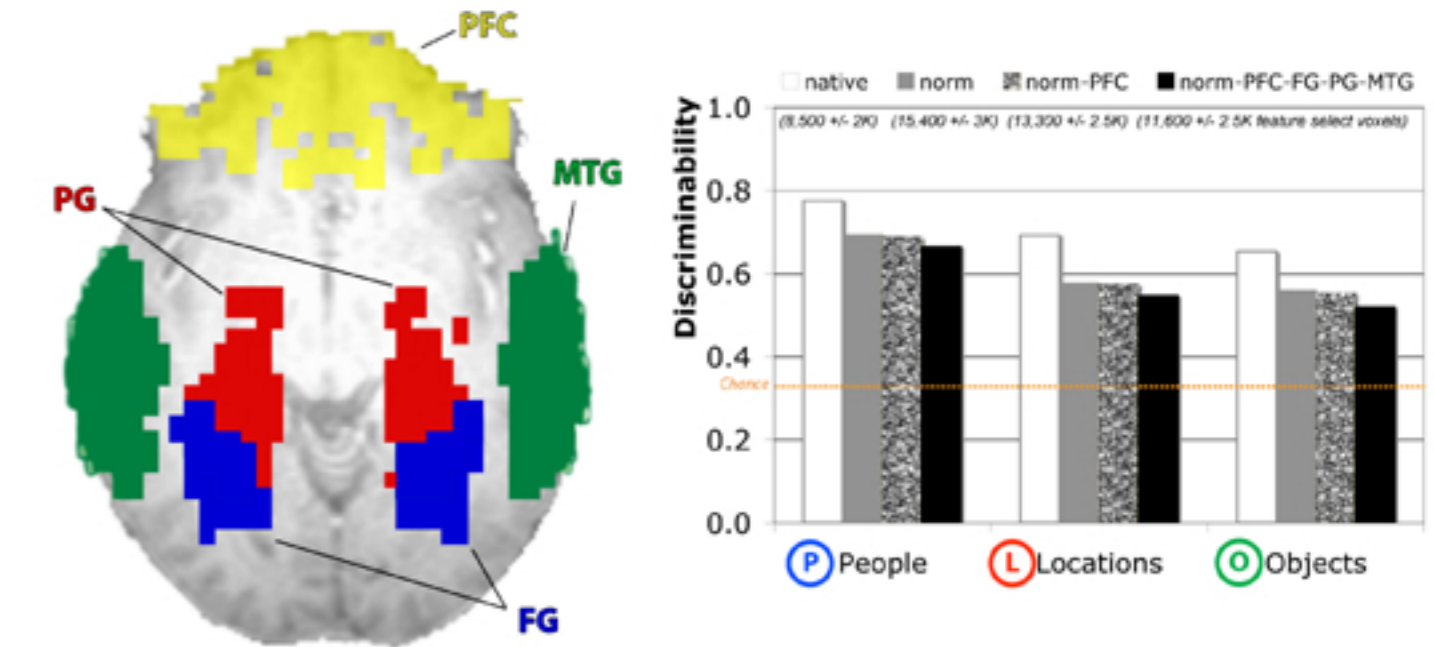
(1 representative subject; native)

- Voxels are colored that were important for discriminating patterns of brain activity corresponding to representations of **People, Locations, & Objects**. (via network weight analysis)
- Canonical category-selective brain areas contributed to the classification of the three categories:

FG, fusiform gyrus; PG, parahippocampal gyrus; MTG, middle temporal gyrus; LO, lateral occipital cortex

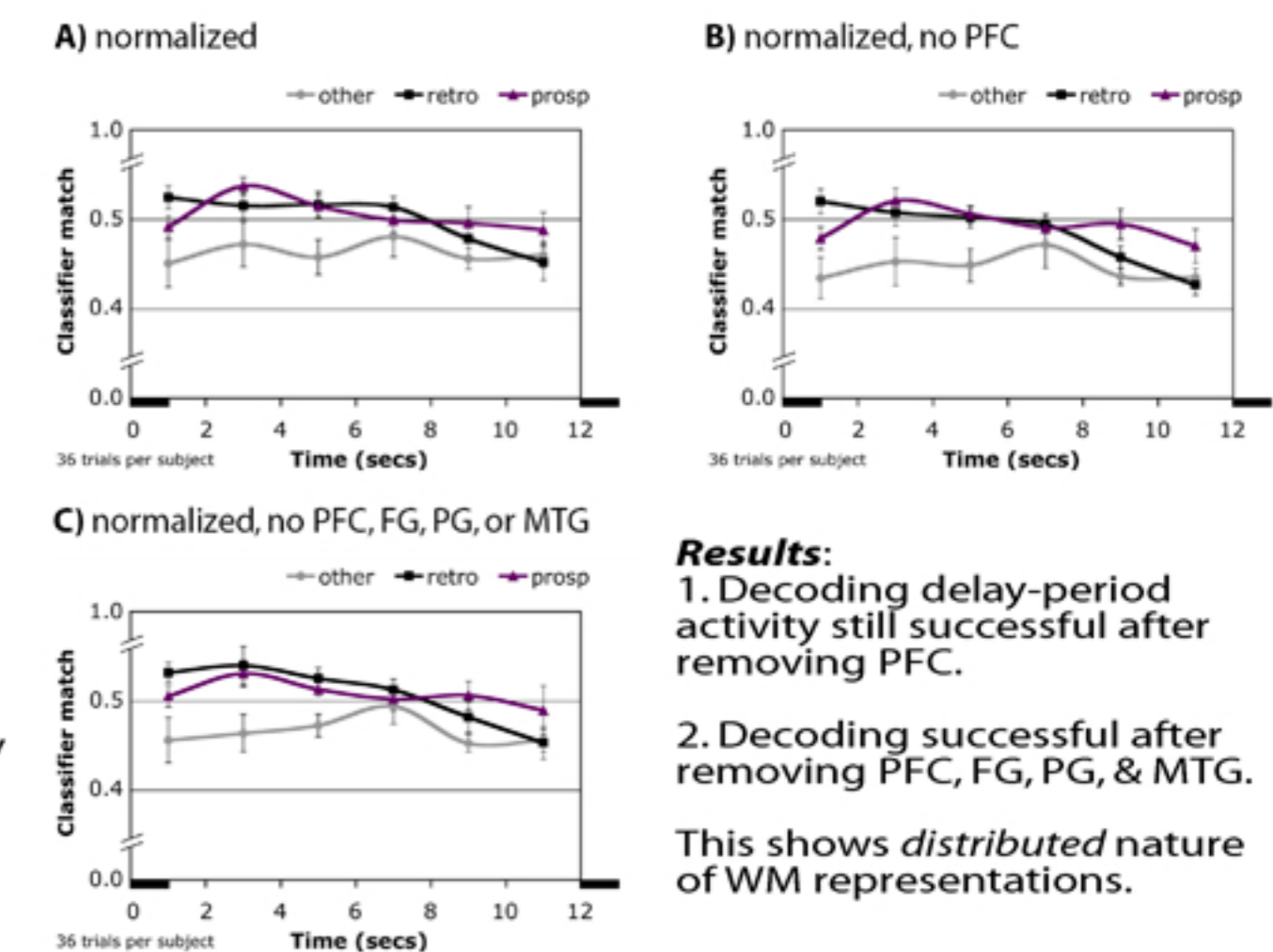


### 3 Are neural reps. localized or distributed?



**Result:** Removing PFC and canonical category-selective regions did not significantly reduce classification accuracy.

This shows *distributed* nature of LTM representations.



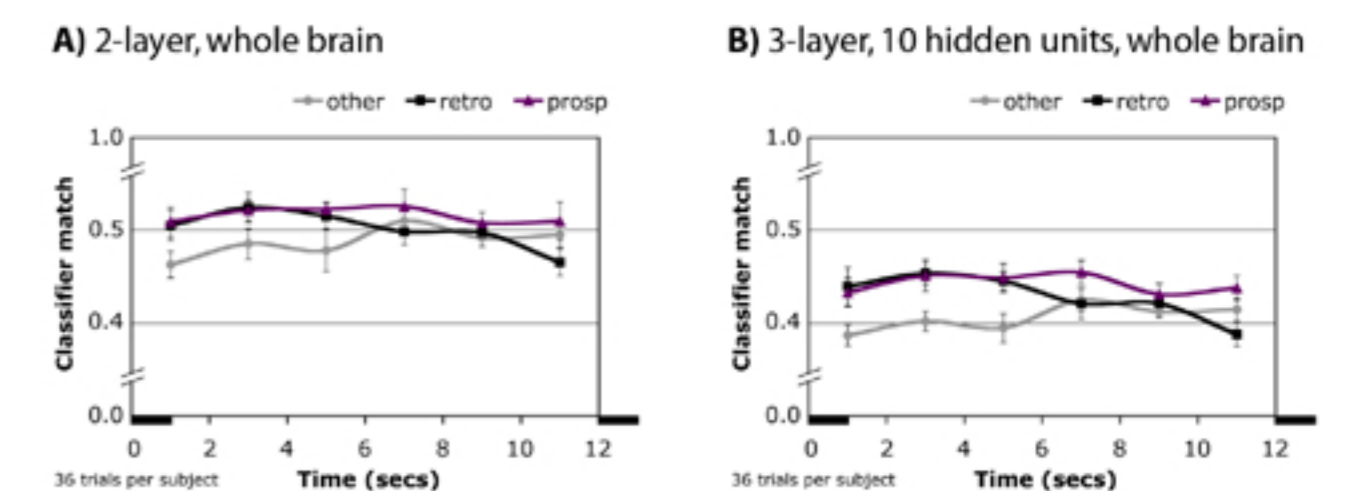
**Results:** 1. Decoding delay-period activity still successful after removing PFC.

2. Decoding successful after removing PFC, FG, PG, & MTG.

This shows *distributed* nature of WM representations.

### 4 Are neural reps. linear or conjunctive?

- Avoided feature-selection (ANOVA,  $p < .05$ ) of voxels. (used voxels from whole brain for classification)
- Added hidden layer to allow for non-linear re-representation.



**Result:** No substantial change in classification performance, implies that voxel activity does not vary across conditions based on activity in other voxels.