# Medial prefrontal cortex compresses concept representations through learning

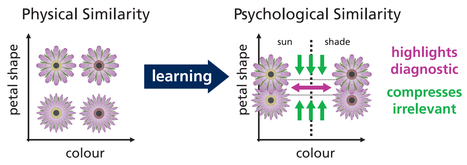

Michael L. Mack[1], {Alison R. Preston[2], Bradley C. Love[3]}*

[1]University of Toronto, [2]The University of Texas at Austin, [3]University College London, *Equal contribution


## Introduction

Prefrontal cortex (PFC) is thought to focus on goal-relevant information. Specifically, medial PFC has been shown to code for latent structures of experience in cognitive maps[1,2,3].

Category learning models formalize this process through feature-based attentional tuning and knowledge clustering[4,5].




This theoretical convergence suggests mPFC may perform data reduction, compressing task-irrelevant features and emphasizing goal-relevant information structures.

## Research Questions

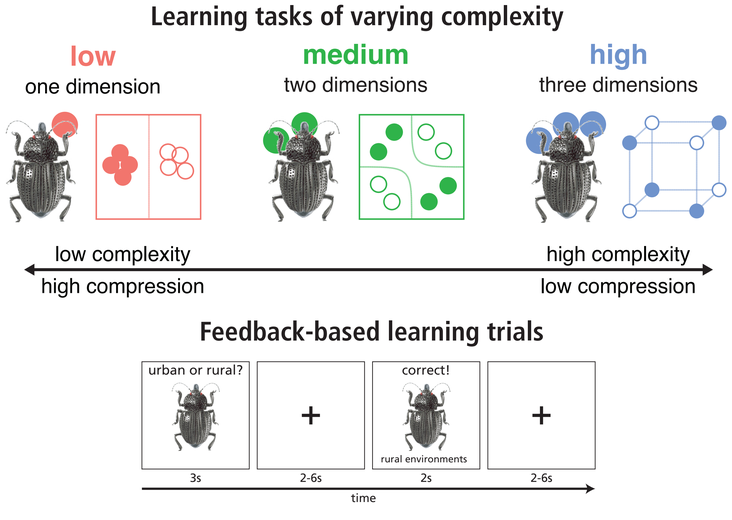
- Does mPFC perform goal-sensitive data reduction during learning?
- Are individual differences in learning ability and attentional tuning predicted by mPFC neural compression?

## Methods

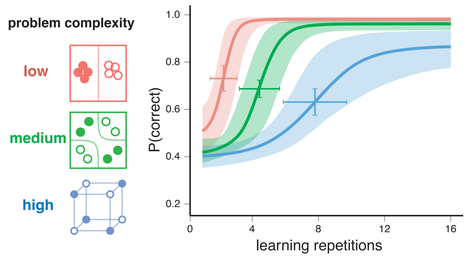
**fMRI category learning tasks** (1.7mm-iso voxels, 2s TR, whole brain)

22 participants learned three classic categorization tasks[6]

4 fMRI runs per task, 32 trials per run, 128 trials total per task

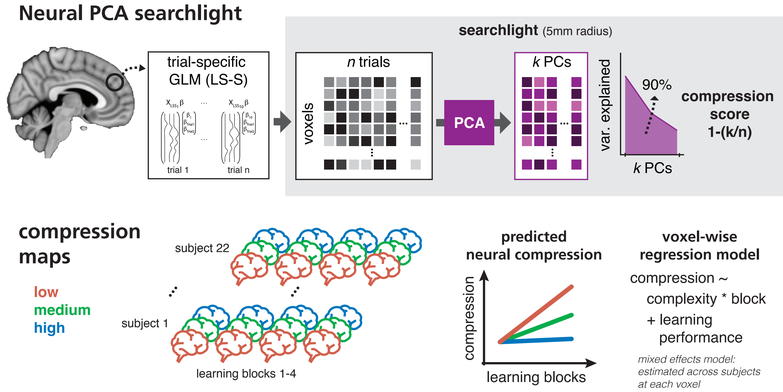



## Learning Behaviour



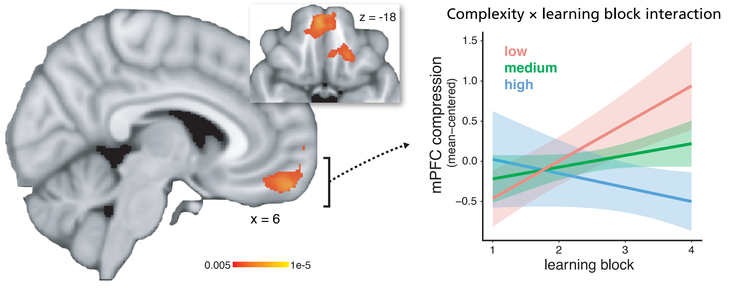

- Both low and medium complexity problems learned quickly
- Complexity related to learning difficulty
- RTs showed corresponding speed-ups with learning

## Indexing neural dimensionality reduction with principal component analysis (PCA)




## Learning-related neural compression

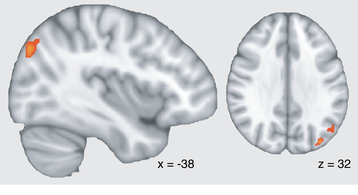
**Medial PFC representational compression during learning**




- mPFC representations show task-specific compression (low > medium > high complexity)
- Neural compression emerges over learning
- Similar pattern in left parietal, but subthreshold

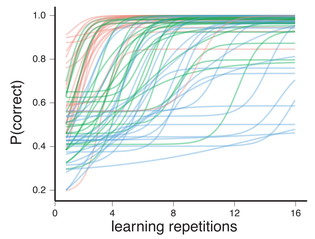
**Left Inferior Parietal**

Cluster extent 167 voxels, P = 0.05 threshold of 176 voxels

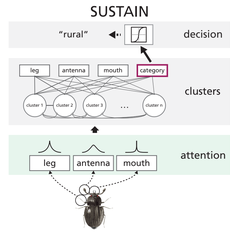


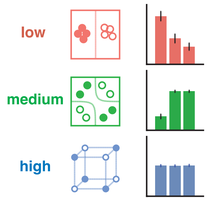

## Neural compression predicts individual differences in learning ability

**Variability in learning ability**




**Computational modeling of learning**

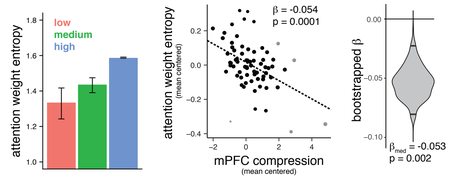



**Predicting attentional tuning with neural compression**

specificity of attentional tuning

$$entropy(w) = -\sum w_i \log(w_i)$$




Greater mPFC compression associated with more optimal task-specific feature attention

## Conclusions

- mPFC representations are shaped by experience through goal-sensitive dimensionality reduction
- mPFC compression consistent with mechanisms of SUSTAIN learning model[4,5], providing quantitative account of mPFC's role in concept *formation*[7]
- Dimensionality reduction through selective attention may underlie many mPFC functions (e.g., cognitive maps, latent causal models, schemas, value coding)


### Acknowledgements

This project is supported by an NSERC Discovery Grant RGPIN-2017-06753, NIMH grants R01-MH100121 and F32-MH100904, NSF CAREER Award 1056019, Leverhulme Trust RPG-2014-075, Wellcome Trust Senior Investigator Award WT106931MA, and NICHD Grant 1P01HD080679. Thank you to Meg Schlichting and Christiane Ahlheim for helpful comments.


### References


1) Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. Neuron, 81(2), 267–278.

2) Schuck, N. W., Cai, M. B., Wilson, R. C., Niv, Y.. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. Neuron, 91(6), 1402–1412.

3) Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nature Communications, 6, 8151.

4) Love, B.C., Medin, D.L., & Gureckis, T.M. (2004). SUSTAIN: A network model of category learning. Psychological Review, 111, 309-332.

5) Love, B.C., & Gureckis, T.M. (2007). Models in search of a brain. Cognitive, Affective, & Behavioral Neuroscience,7, 90-108.

6) Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. Psychological Monographs, 75.

7) Mack, M. L., Love, B. C., & Preston, A. R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. Proceedings of the National Academy of Sciences, 113(46), 13203-13208.



NSERC CRSNG

MACK LAB

mack@psych.utoronto.ca

http://macklab.utoronto.ca