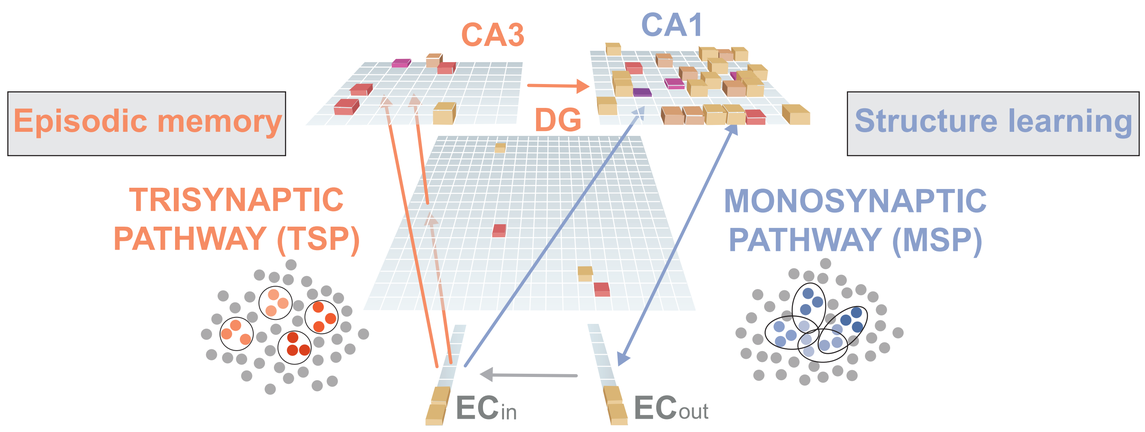
A mixture-of-experts model of hippocampal-mPFC interactions

Dhairyya Singh, Ashley B. Williams, Anna C. Schapiro
Department of Psychology, University of Pennsylvania

Penn Computational Cognitive Neuroscience Lab

Introduction

The hippocampus is critical for both episodic memory and structure learning, functions that benefit from opposing styles of representation.
C-HORSE[1] proposes that the hippocampus (HPC) contains an anatomical dissociation to support these functions:

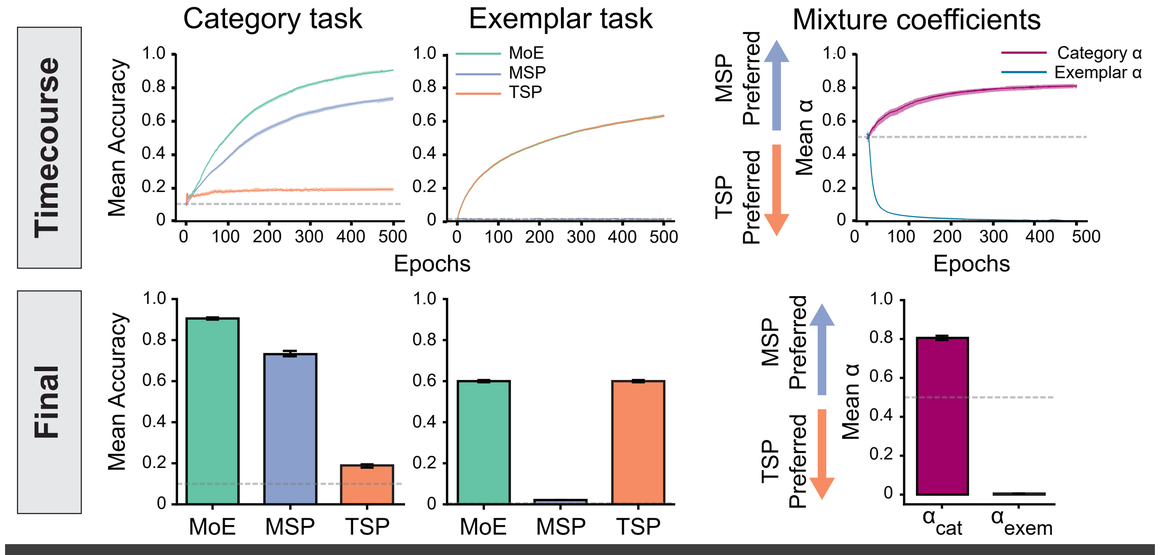


How does the brain deploy the most useful representation for the current task?
Medial prefrontal cortex (mPFC) has been implicated in exerting control over hippocampal function[2] and could play a role in deploying task-adaptive representations.

We provide an account of these interactions using a stimulus-computable **mixture-of-experts system (MoE)[3], where an mPFC-inspired output gating network learns to control the task-specific deployment of hippocampal representations.**

Results

Expert performance & gating

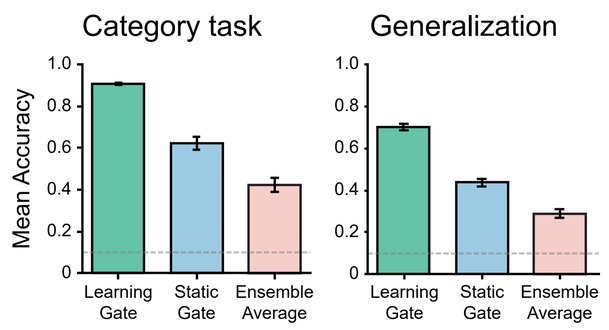


Experts exhibit a double dissociation in specialization, which is effectively leveraged by gating network.

Examining the importance of learned gating

Static gate: equal α (0.5) for experts throughout training.

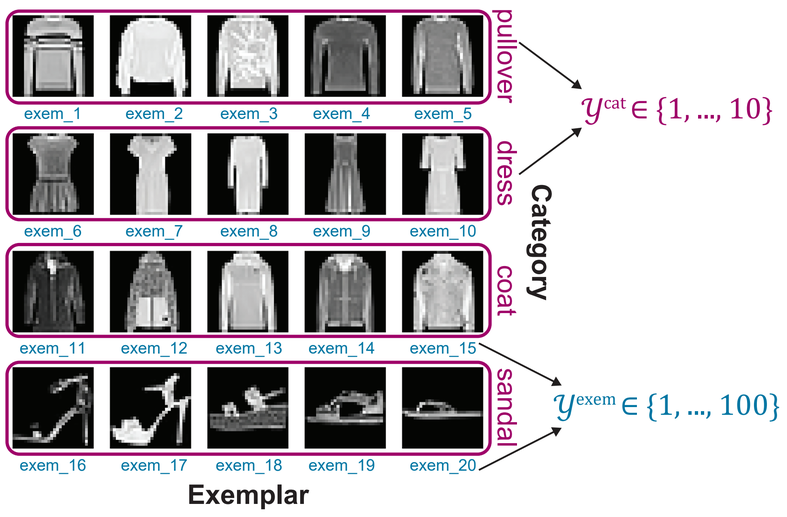
Ensemble average: independently trained experts averaged at test.



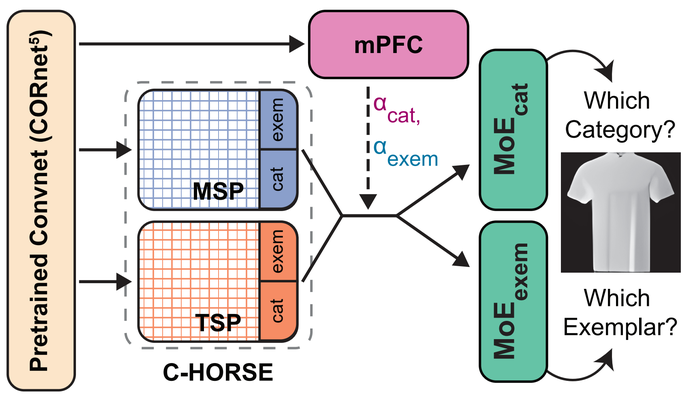
Learned gating is critical for robust model performance.

Architecture & training

Fashion-MNIST[4] subset (N=100; $\mathcal{X}$)



Mixture-of-Experts model architecture



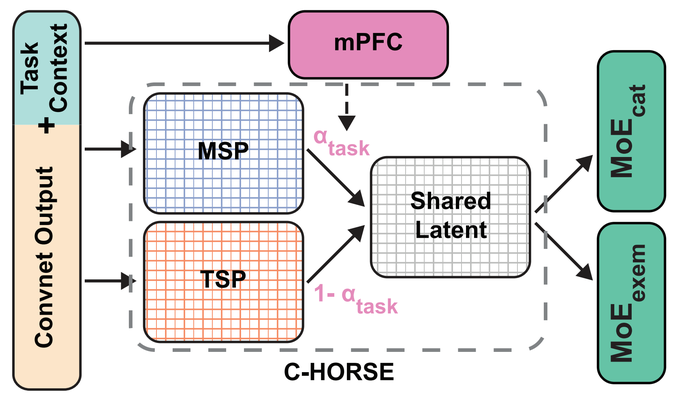
$$\alpha_{task} \in [0,1]; \mathrm{MoE}_{task} = \alpha_{task} * \mathrm{MSP}_{task} + (1 - \alpha_{task}) * \mathrm{TSP}_{task}$$

	MSP	TSP	mPFC
Hidden layer size	Small	Large	Small
Learning rate	Low	High	Low
Sparsity/Inhibition	Nil	High	Nil

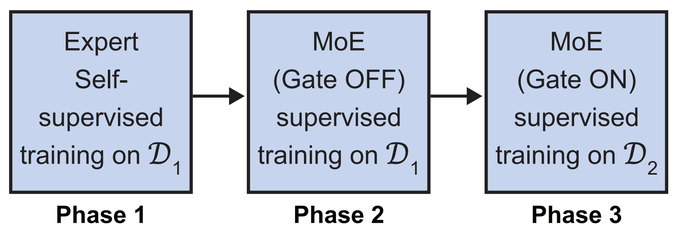
Current directions

MSP and TSP representations **converge in HPC subfield CA1,** a region **directly targetted and modulated by mPFC.** This modulation influences **both generalization and discrimination behavior[6,7].**
To explore **how mPFC shapes CA1 representations,** we extend the model to include a shared latent space. We also examine how **prior learning shapes expert generalization and gating dynamics** via a blocked curriculum:

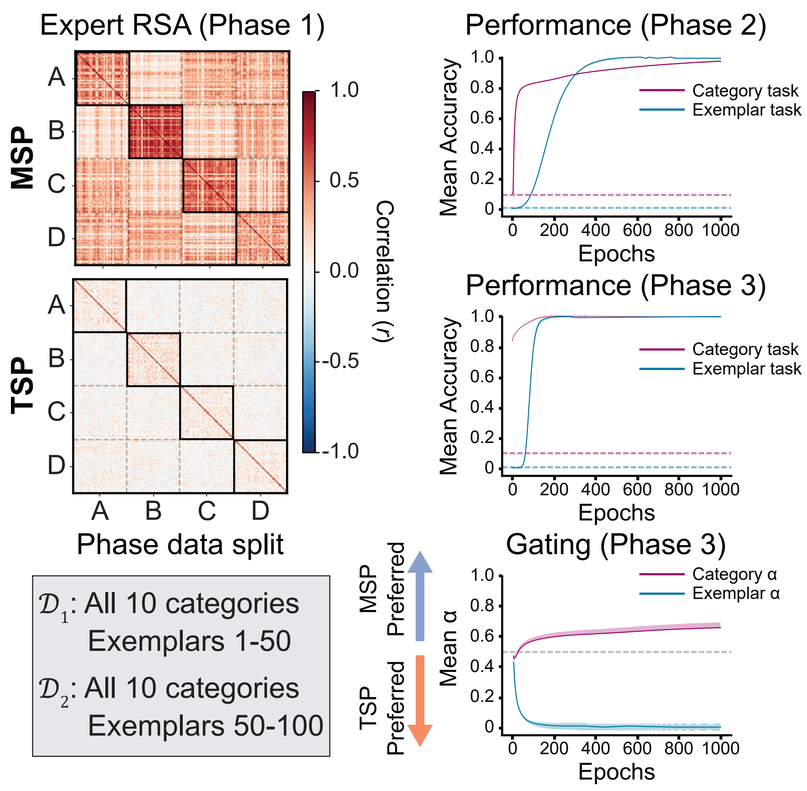
Extended model architecture & training



Training curriculum



Extended model results



$\mathcal{D}_1$: All 10 categories Exemplars 1-50
$\mathcal{D}_2$: All 10 categories Exemplars 50-100

Self-supervised learning promotes representational differentiation and both experts and the gating network **leverage prior learning** for enhanced performance.
Future work will examine the shared latent representations and target a range of experimental findings on HPC-mPFC interactions in rodents and humans.

Take Home

Our mixture-of-experts model demonstrates (i) how mPFC can flexibly leverage complementary hippocampal representations for task appropriate memory use and (ii) that learned gating is critical for robust performance.
In ongoing work we are exploring the role of mPFC in shaping converging MSP and TSP representations in CA1, and how prior learning impacts expert specialization and gating dynamics.

References: 1. Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017), *Philos. Trans. R. Soc. B.* 2. Eichenbaum, H. (2017), *Nat. Rev. Neurosci.* 3. Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991), *Neural Comput.* 4. Xiao, H., Rasul, K. & Vollgraf, R. (2017), *arXiv*. 5. Kubilius, J., Schrimpf, M., Kar, K., Rajalingham, R., Hong, H., Majaj, N., Issa, E., Bashivan, P., Prescott-Roy, J., Schmidt, K., Nayebi, A., Bear, D., Yamins, D. L., & DiCarlo, J. J. (2019). *NeurIPS.* 6. Xu, W & Südhof T.C. (2013), *Science*. 7. Guise, K. G. & Shapiro, M. L. (2017), *Neuron*.

@dhairyya.singh.bsky.social dsin@sas.upenn.edu