Posterior distributions over hidden variables: Schemas in the brain

Stephanie C.Y. Chan Yael Niv Kenneth A. Norman

PRINCETON UNIVERSITY

PRINCETON Neuroscience INSTITUTE

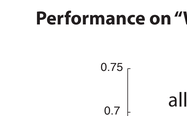
1 Introduction

Situation models and "schemas"

- Serve to organize thoughts and experiences as we encode them into memory
- The posterior-medial network (PM network) of brain regions might be involved in their construction and application

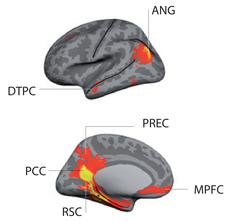
Central question: How do we infer what situation we are in?

Using **Bayesian latent cause models?**



- Situations can be viewed as hidden causes that give rise to observable events
- We can use Bayesian inference to infer the current situation, as the **posterior probability distribution** ***P(situation | observations)***

The posterior-medial network

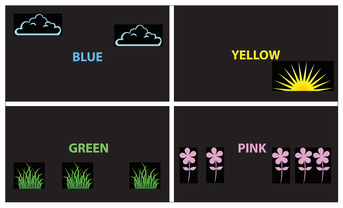


(Ranganath & Ritchey 2012)

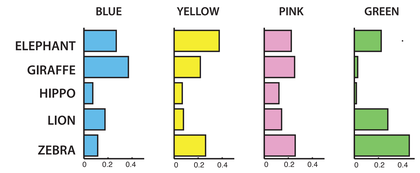
Hypothesis: Brain regions implicated in situation modelling (the PM network) represent the posterior distribution over situations, as computed by Bayesian latent cause models.

2 "Animal Safari" posteriors task

The safari is divided into 4 "zones"

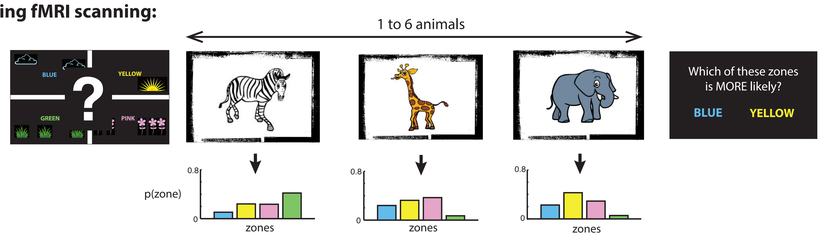


Animals appear in different zones with different probabilities:



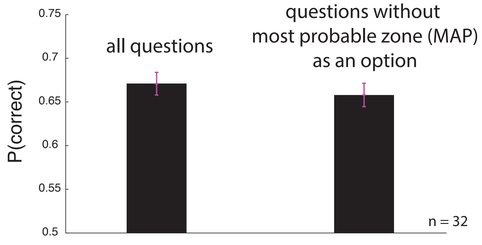
Subjects are trained on these probabilities before scanning.

During fMRI scanning:



Subjects must continuously update their beliefs about the **posterior probability of each zone.**

Performance on "Which zone is more/less likely":



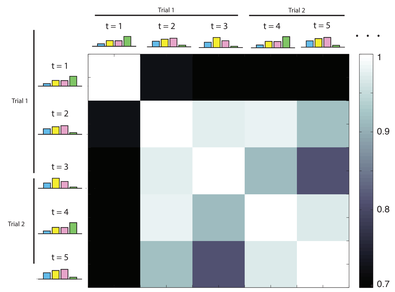
- Subjects perform significantly above chance
- Subjects are not just representing "which zone is most likely"

3 Which areas represent the posterior?

Representational similarity analysis (RSA)

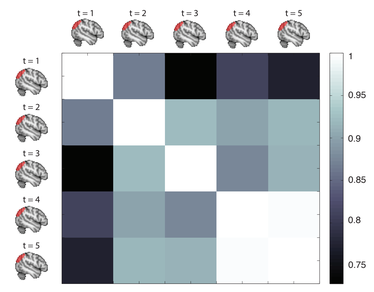
Or alternate models

Similarity structure for **posterior** [T x T]



What the similarity structure *should* look like

Neural similarity structure for **ROI** [T x T]

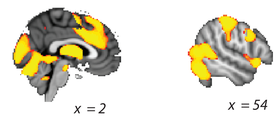


What the similarity structure *actually* looks like

Take correlation of these two matrices to obtain the **representational similarity match of ROI and posterior**

Or alternate models

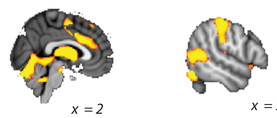
match to posterior



n = 27
(excluding subjects with poor performance)

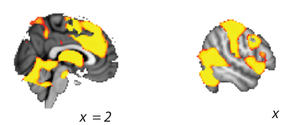
p < 0.05
FWE corrected

match to posterior > match to MAP



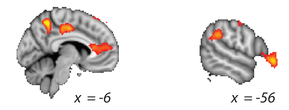
MAP = maximum a posteriori
i.e. the most probable sector

match to posterior > match to current animal



current animal
= which of the five animals are currently being shown

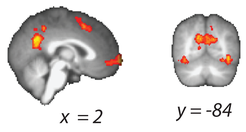
match to posterior correlates with behavioral performance (across subjects)



n = 32
p < 0.05 uncorrected

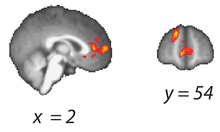
4 Univariate analyses

Which areas *update* the posterior?



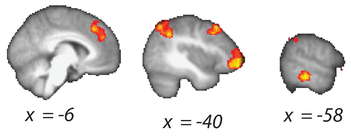
parametric regressor:
KLdiv(Posterior at t, Posterior at t-1)

Which areas represent surprise?



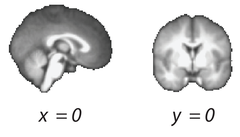
parametric regressor:
P(current animal | animals seen so far)

Which areas represent confidence?



parametric regressor: P(MAP)

Which areas represent difficulty?



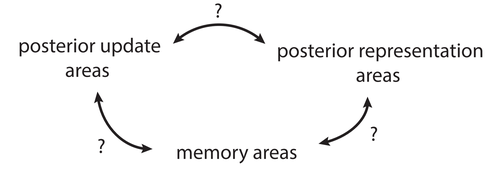
parametric regressor: entropy

5 Next steps

Further alternate models for RSA

- Difficulty / attention / uncertainty / conflict
- Associative / Hebbian model
- Reinforcement learning / temporal-difference model

Connectivity analyses



Relationship with behavior

- Trial-by-trial correlations
- Try to infer likelihood and posterior representations by modeling behavior

Acknowledgements

This research was supported by the National Institute of Mental Health (award T32MH065214); the NSF Collaborative Research in Computational Neuroscience (CRCNS) Program (award number IIS1009452), and the U. S. Army Research Laboratory and the U. S. Army Research Office (contract/grant number W911NF1410101).