Neural Mechanism of Compositional Generalization

Zilu Liang[1], Leonie Glitz[1,2], Miriam Klein-Flugge[1], and Christopher Summerfield[1]
[1] Department of Experimental Psychology, University of Oxford, Oxford, UK
[2] Medical Research Council Brain Network Dynamics Unit, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

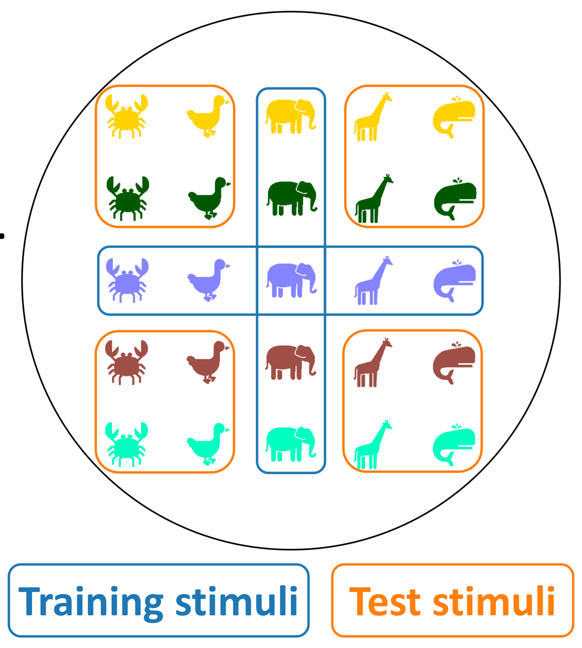
Introduction

- Humans learn building blocks that can be recomposed to solve novel problems efficiently.
- Using an established compositional rule generalization task (Dekker et al., 2022) and function MRI, we studied the neural representations in hippocampus (HPC), ventral medial prefrontal cortex (vmPFC), posterior parietal cortex (PPC), and primary visual cortex (V1) in generalizers.
- We investigated: 1) the neural geometry of rule representations and 2) the composition of rule representations that support generalization

The treasure hunt task

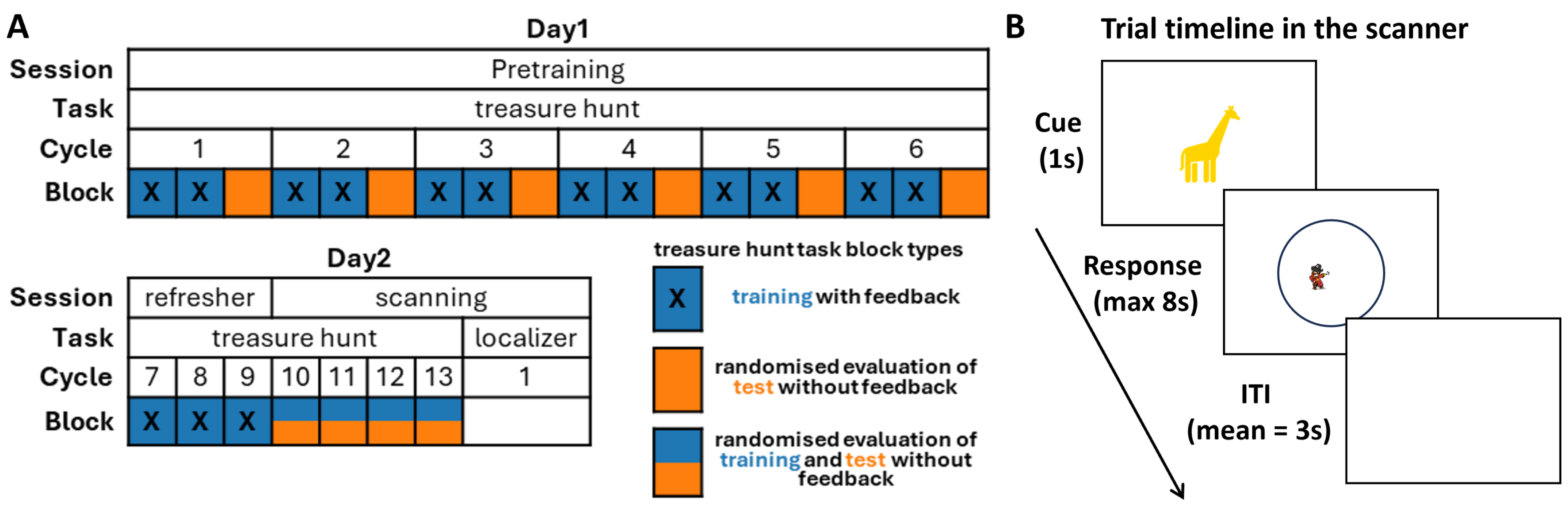
Design

- Task: predict treasure locations based on cue stimuli
- Rules: discrete colours/shapes define x-/y-axis locations.
- Participants only received feedback about the correct locations for the training stimuli.
- Generalization performance were evaluated with test stimuli without feedback.



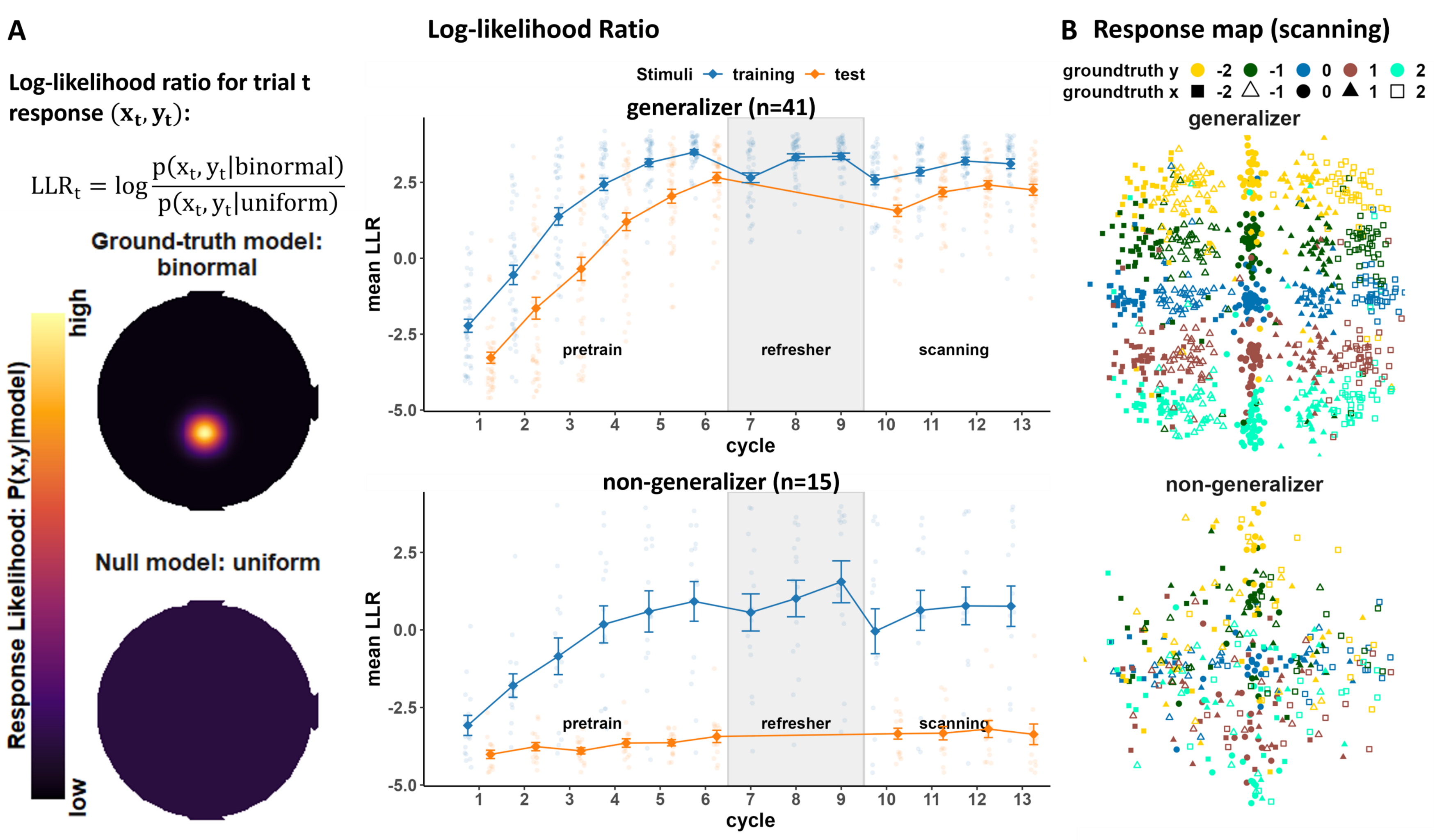
Procedure

- Participants completed pre-training at home in Day1.
- Participants were scanned while performing the treasure hunt task in Day2.



Generalization Performance and Participant Classification

- Following Dekker et al., 2022, participants were classified into generalizers and non-generalizers based on their performance in the test trials.
- 41/56 participants were generalizers.
- **We reported the neural results in generalizers**



Acknowledgements

This research was generously funded by a Wellcome Trust Discovery Award (227928/Z/23/Z) to C.S. M.C.K.-F. is funded by a Wellcome Henry Dale Fellowship (223263/Z/21/Z) and a UKRI-underwritten ERC Starting Grant (EP/X021815/1).

If you have any questions, feel free to contact zilu.liang@psy.ox.ac.uk

You can also find the online version of this poster and the abstract by scanning the QR codes:

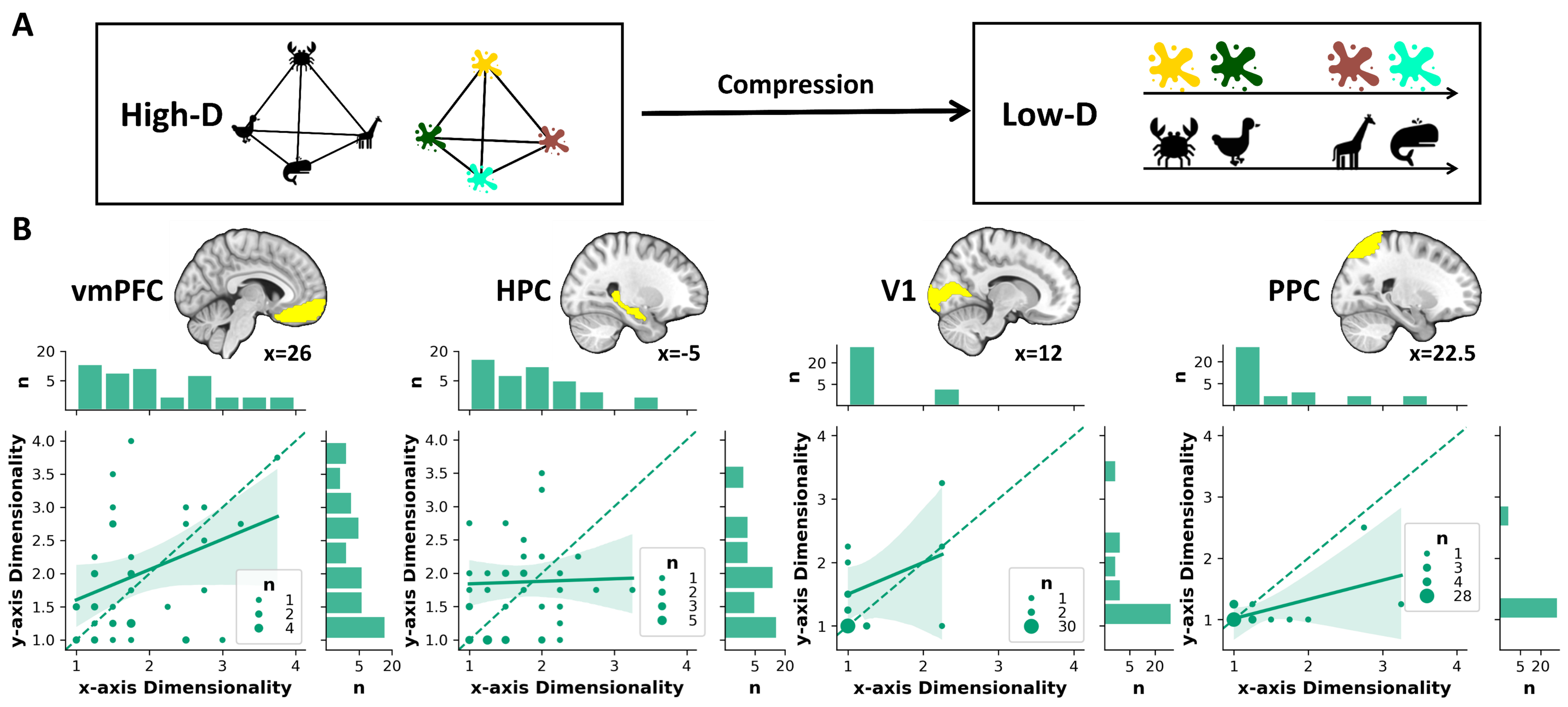
Abstract Poster Task Demo Pre-registration

Geometry of rule representations

Dimensionality

Compression of high-dimensional feature representations into low-dimensional axes?

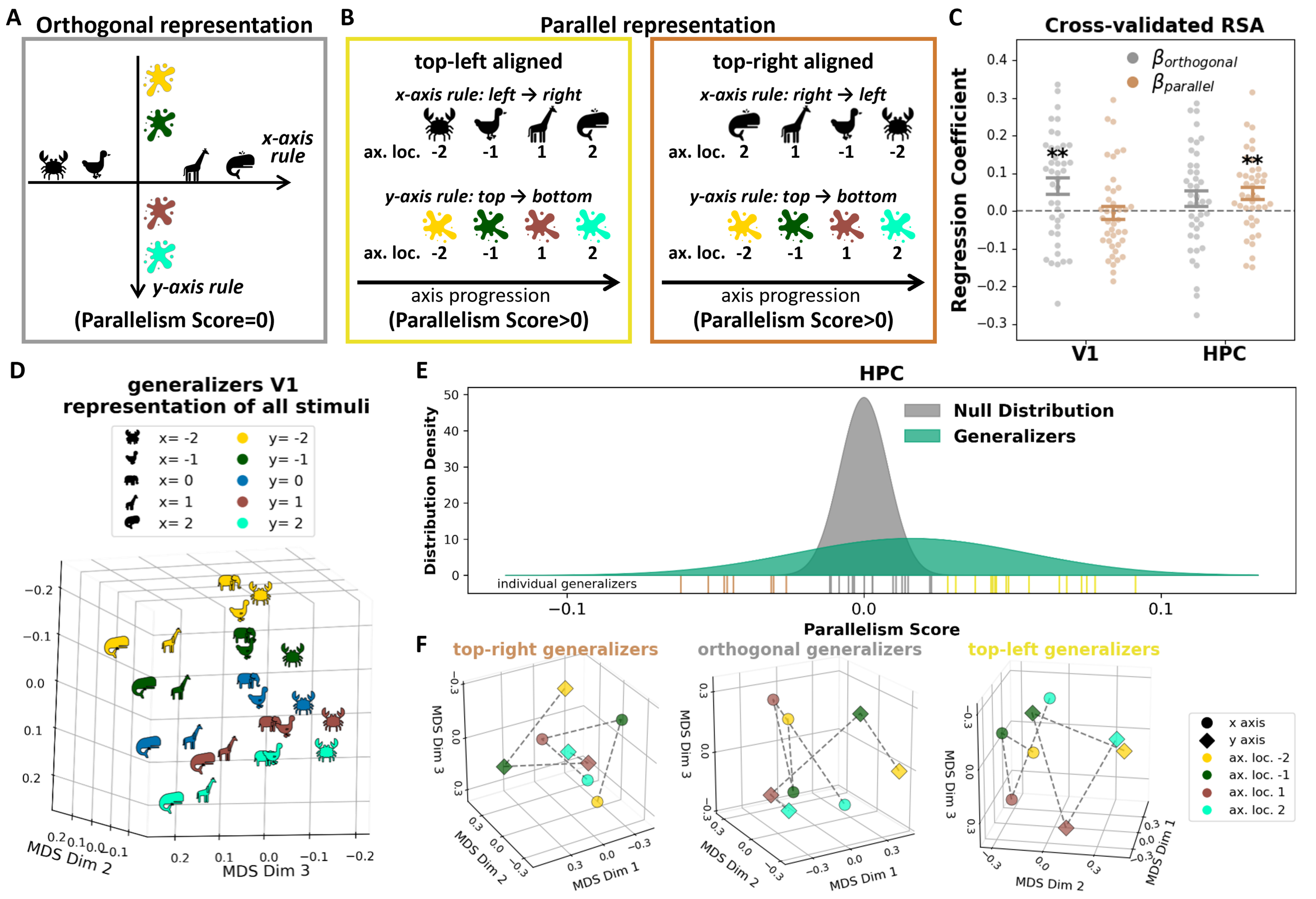
HPC and vmPFC maintained high-dimensional, PPC and V1 became low-dimensional



Orthogonality

Relationship between representations of different rules:

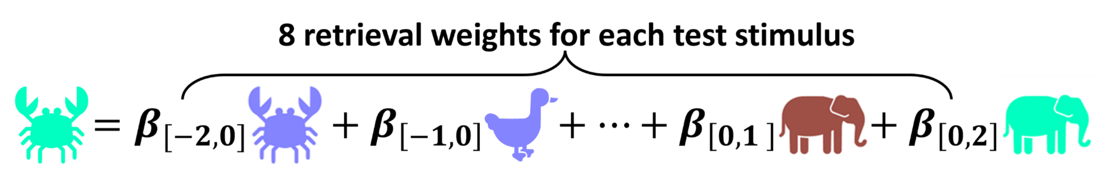
Representations of rules were orthogonal in V1, but parallel in HPC



Neural composition via vector addition

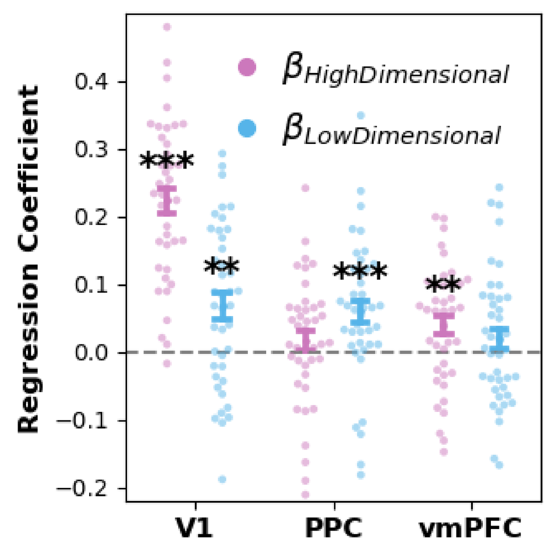
Composing test stimuli representations from rule representations

- Predict the representations of test stimuli as weighted sum of training stimuli representations.



- Fitted weights were evaluated in held-out data. $R^2>0$ indicates ssuccessful prediction. **Only V1, PPC and vmPFC showed vector addition composition.**

High- vs low-dimensional composition



- Regressed the retrieval weights on two models:
  - High-D: retrieval weights modulated by *feature relevance*
  - Low-D: retrieval weights modulated by *spatial distance*
- Consistent with the dimensionality results:
  - **vmPFC was high-dimensional**
  - **PPC was low-dimensional**
  - **V1 represented features as well as spatial locations**

Discussion and Future directions

Summary

- **We observed variations in dimensionality and orthogonality of rule representations, which is likely related to the functional roles of different regions when performing the task.**
- **We validated empirically that vector addition support the composition of independent rule components when generalizing to novel stimuli.**

Next steps

- Modelling the variations in parallel vs orthogonal rule representation
- Cross-context compositional generalization using a multi-map design