Generalization of Covariance Structure in Human and Neural Network

Zilu Liang, Miriam Klein-Flugge, Christopher Summerfield
Department of Experimental Psychology, University of Oxford

Introduction

- A task on learning covariance structure among feature variables and transfer to 1) a new response space, and 2) new combinations of learned features.
- Investigate effective curriculum for structural learning and transfer in humans.
- Derive a theoretical model of representation that support structure transfer in the task and test predictions with neural network simulations

The label prediction task (9-grid)

- constructed 27 stimuli from the factorial combination of 3 shapes, 3 colours, and 3 textures
- Participants learn to predict labels of the stimuli based on these features.
- **Task rule - One pair of correlated colours:** *blue and yellow stimuli of the same shape always have the same label.*

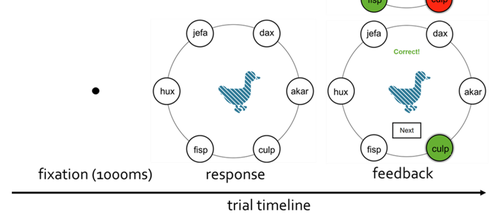
train			
Whale	akar	akar	fisp
Elephant	dax	dax	culp
Duck	hux	hux	jefa

transfer			
Whale	rel	rel	erag
Elephant	kern	kern	gip
Duck	lep	lep	blap

trial type: training (with feedback) | anchor | test (no feedback)

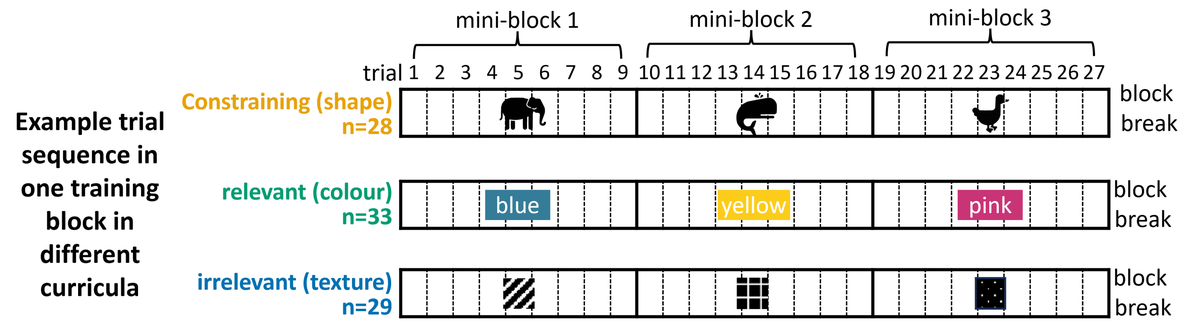
- Training: mapping stimuli to a set of 6 labels.
- Transfer: mapping stimuli to a new set of 6 labels.
- 10 training-transfer cycles (Interleaved train/transfer blocks)

Timeline of trial with feedback

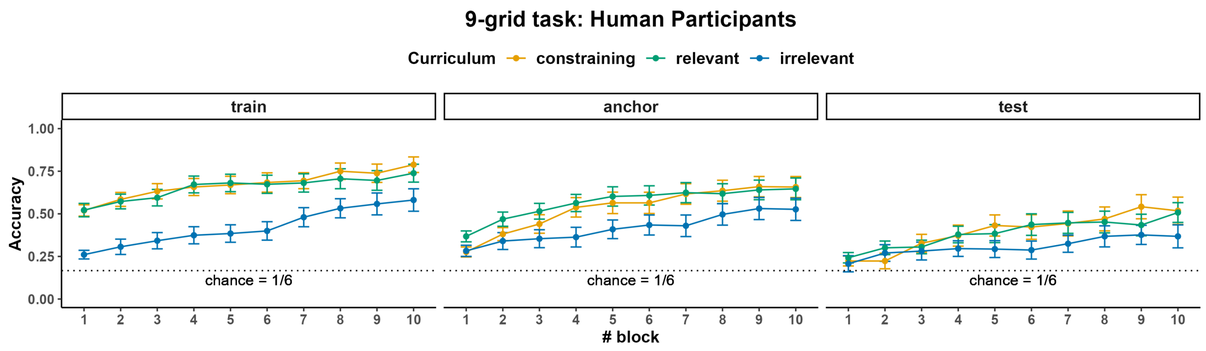


Human participants performance under different training curricula

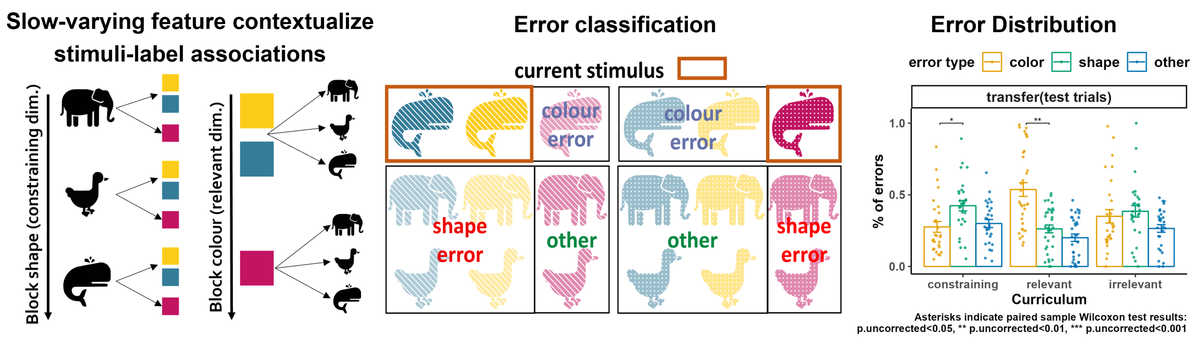
- We set up three training curricula that introduce temporal autocorrelation (blocking) to 1 of the 3 dimensions.



- Participants are randomly assigned to one of the three training curricula.



- The slow varying feature in the training stimuli contextualize stimuli-label associations
- Participants are less likely to be confused by associations from the same context than to be confused by associations from different contexts (Collins & Koechlin, 2013).



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:

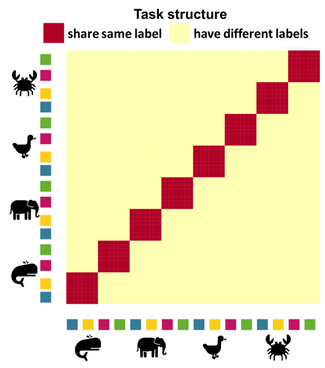
Abstract

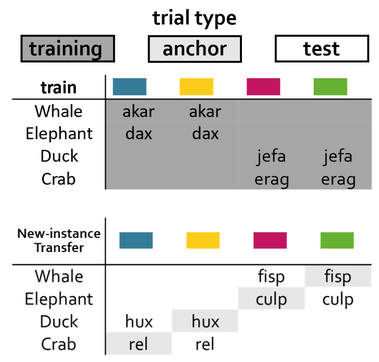
Poster

Extension to 16-grid label prediction task and neural network simulation

Extension to 16-grid task

- 64 stimuli generated from 4 shapes, 4 colours and 4 textures
- **two pairs of correlated colours:** *blue-yellow and red-green*





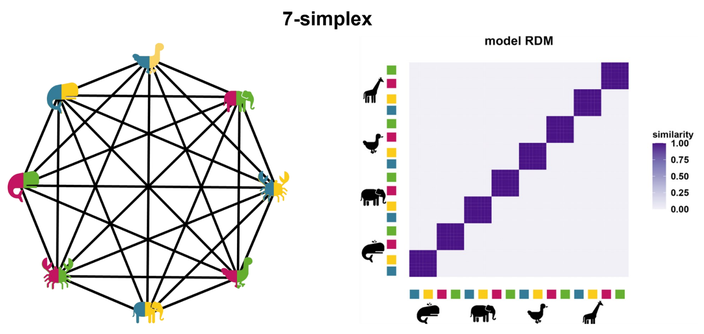
Neural Network Simulation

- Architecture: MLP with one hidden layer trained with SGD update
- A full training procedure as behavioural study and two control training procedures disrupting learning

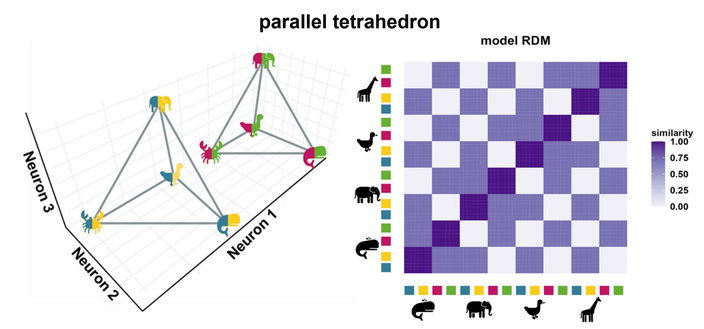
training procedures	trial type: training	anchor	Test
full	√	√	√
control 1: anchor only	X	√	√
control 2: train random	Train on random stimuli-label mapping	√	√

Hidden layer representation that supports structure transfer in neural network

Theoretical models of representation geometry



- **7-simplex model** : A high dimensional representation hinders the generalization of the relational structure.



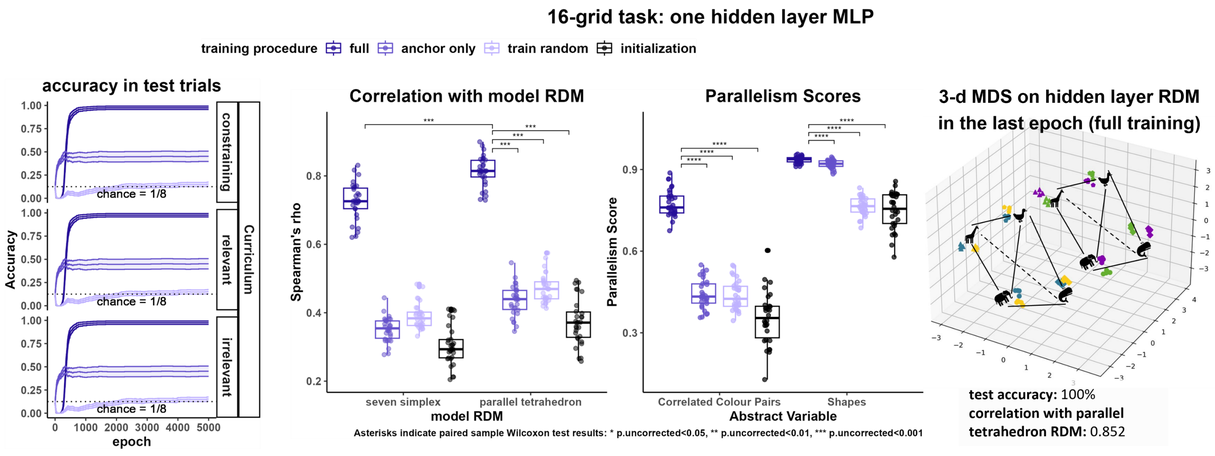
- **Parallel tetrahedron model**: A low dimensional abstract representation that reproduces the latent structure supports knowledge generalization (Bernardi et al., 2020; Johnston & Fusi, 2022).

Quantifying abstraction in hidden layer representation

- Correlation between hidden layer RDM and model RDM
- Parallelism Score (Bernardi et al., 2020; Ito et al., 2022) of abstract variables: correlated colour pairs and shape

Result

- MLP with one hidden layer trained with SGD update is not sensitive to training curriculum
- After full training , the network learns the structure and can transfer
- After full training, hidden layer representation resembles the parallel tetrahedron model



Discussion and Future directions

- Structure learning and transfer in human participants is sensitive to training curriculum
- Neural geometry that represents the stimuli in a low-dimensional space spanned by a set of latent variables supports zero-shot generalization of relational structure to new compositions of stimuli features

Next steps

- Modelling curriculum effect
- Behavioral study of the 16-grid task

Acknowledgements

This work was supported by the European Research Council (award REP-725937 to C.S.) and Wellcome Henry Dale fellowship (223263/Z/21/Z to M.K.F.).