Curriculum Effects in Compositional and Asynchronous Generalisation

Carla Zoe Cremer   Jacques Pesnot Lerousseau   Christopher Summerfield

University of Oxford, Department of Psychology

Background

Compositional generalisation is the ability to de-compose perceptual features and recognise them in *new* constellations.

We extend [Dekker et al., 2022] to study how **curricula (the statistical structure of a training regime)** affect humans and neural networks in *whether* they generalise a compositional rule and *how* they temporally generalise across multiple stimulus dimensions: **simultaneously or sequentially (asyncronicity)**.

We consider two theories of generalisation: rule-based extrapolation and an exemplar-based approximation of previously seen stimuli.

Hypotheses

Hyp1.I The **blocked curriculum (BC)** *promotes* and the **interleaved curriculum (IC)** *compromises* rule-learning and generalisation. **Hyp1.II** BC promotes asyncronous generalisation.

Experimental Design

Online participants learned to map non-spatial features to 3D spatial locations.

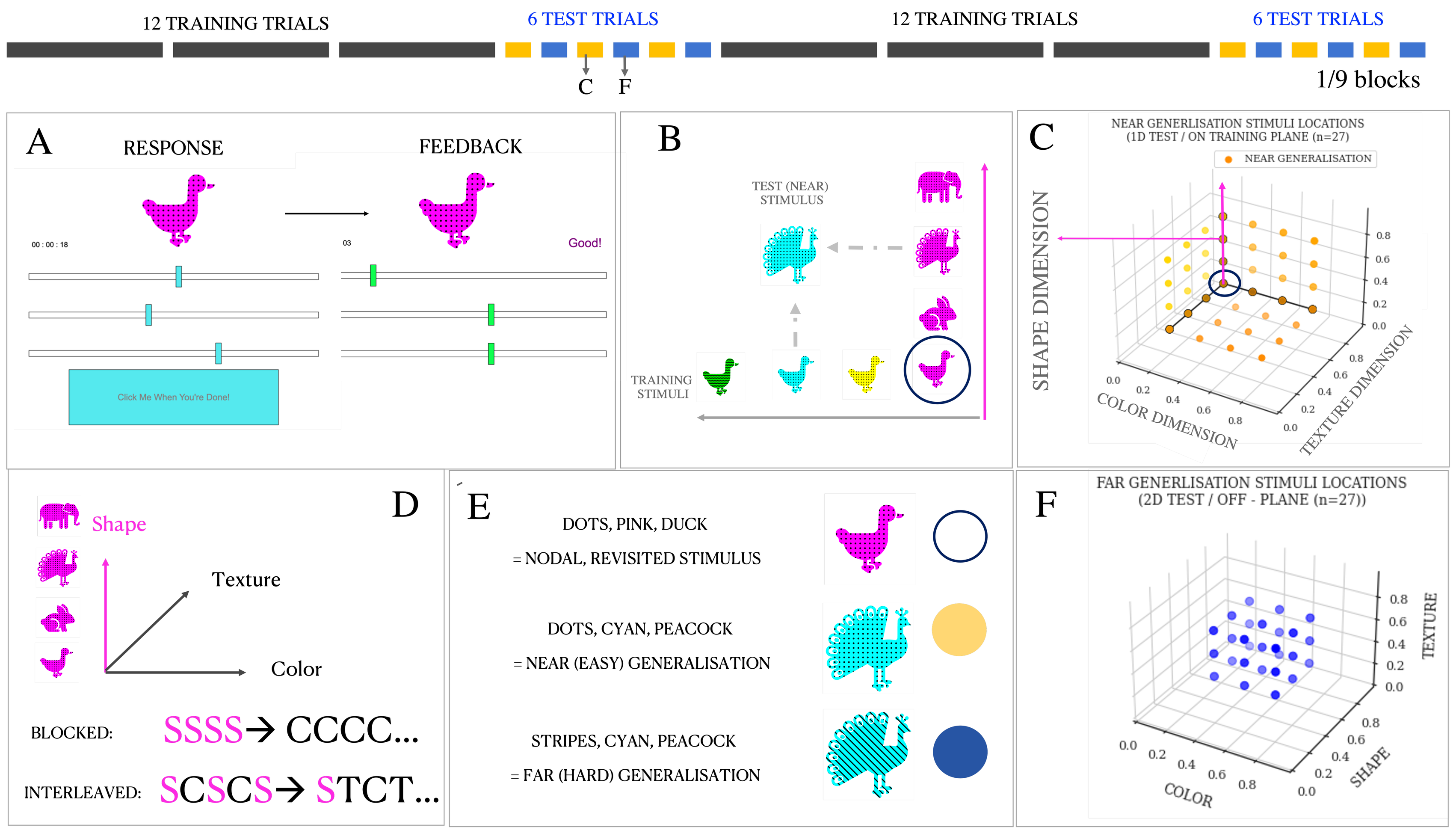

Figure 1. **(A)** Single trial, showing stimulus with 3 features (e.g. pink, duck, dots) and response space per slider. After response, the correct slider locations are shown as supervision. **(B)** Training stimuli along 2 axes (shape and color) with a shared, revisited nodal stimulus (pink, duck, dots). Test stimulus with 1 new feature combination (**near generalisation**). **(C)** Near (yellow) test locations lie on and **far** (blue) test stimuli **(F)** lie off the planes of training axes. **(D)** The **training curriculum is the independent variable**.

Compositional Rule-Learning and Generalisation

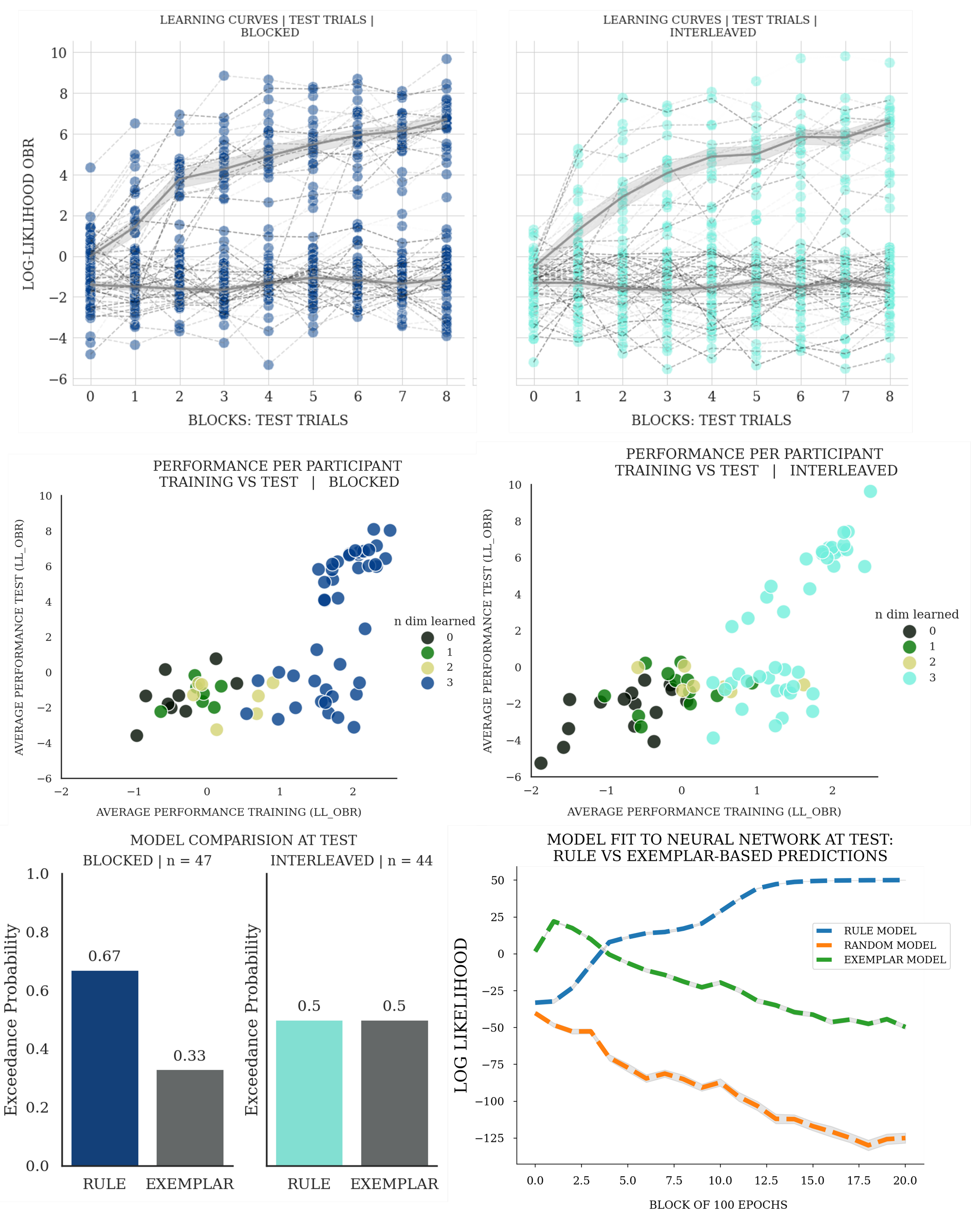

Figure 2. **Top:** Mean log-likelihood of correct response per block for participants (dots) in BC (left) and IC (right) during test. Clear separation between generalisers and those who failed (median in grey). **Middle:** Average performance per participant. Blue (left, blocked) and cyan (right, interl.) signifies that participant learned and/or generalised across all 3 dimensions. **Bottom:** (left) **posterior frequency (post. freq.)** of rule vs exemplar model fit to humans (who were 3D at training) at test (BF=1.04). (right) rule, exemplar, random model fit to single neural network at test. Exemplar-like approximation first increases and then decreases as rule-like predictions are learned. We find no curriculum effects in neural networks.

Models

Our **performance metric** LL_{obr} captures correct, random and biased behavior.

$$LL_{\text{obr}} = LL_{\text{o}} - \max(LL_{\text{b}}, LL_{\text{r}}) \tag{1}$$

Where O = $N(\mu = \text{groundtruth}, \sigma = \text{fit by grid-search})$, B = Gaussian centered on the most frequent response and R = uniform distribution across all locations.

Neural network: linear, 12 one-hot inputs, 3 hidden, 3 output nodes (sigmoid activation), MSE loss

Rule and exemplar-based models: rule-model is $N(\mu = groundtruth, \sigma = 0.1)$ and exemplar-model is a Gaussian likelihood function centered on an average of known stimulus locations that share features with the test stimulus.

Syncronous (SY) and **Asyncronous (AS)** models (sigmoids) are fit to learning curves. In SY model, inflection point is fixed across dimensions. In AS model it is a free parameter.

We conduct a Bayesian model comparison across curricula.

Table 1. Curriculum Effect | Posterior Frequency

		fail (<3D)	success (=3D)	AS
Blocked	Train	0.24	0.76	0.02
	Test	0.59	0.41	0.45
Interleaved	Train	0.41	0.59	0.19
	Test	0.71	0.29	0.29

Asyncronicity

We find **asyncronicity at test** in humans and neural networks. Blocking lead to (insignificantly, BF = .32) more asyncronicity (during test) than interleaving (Table 1).

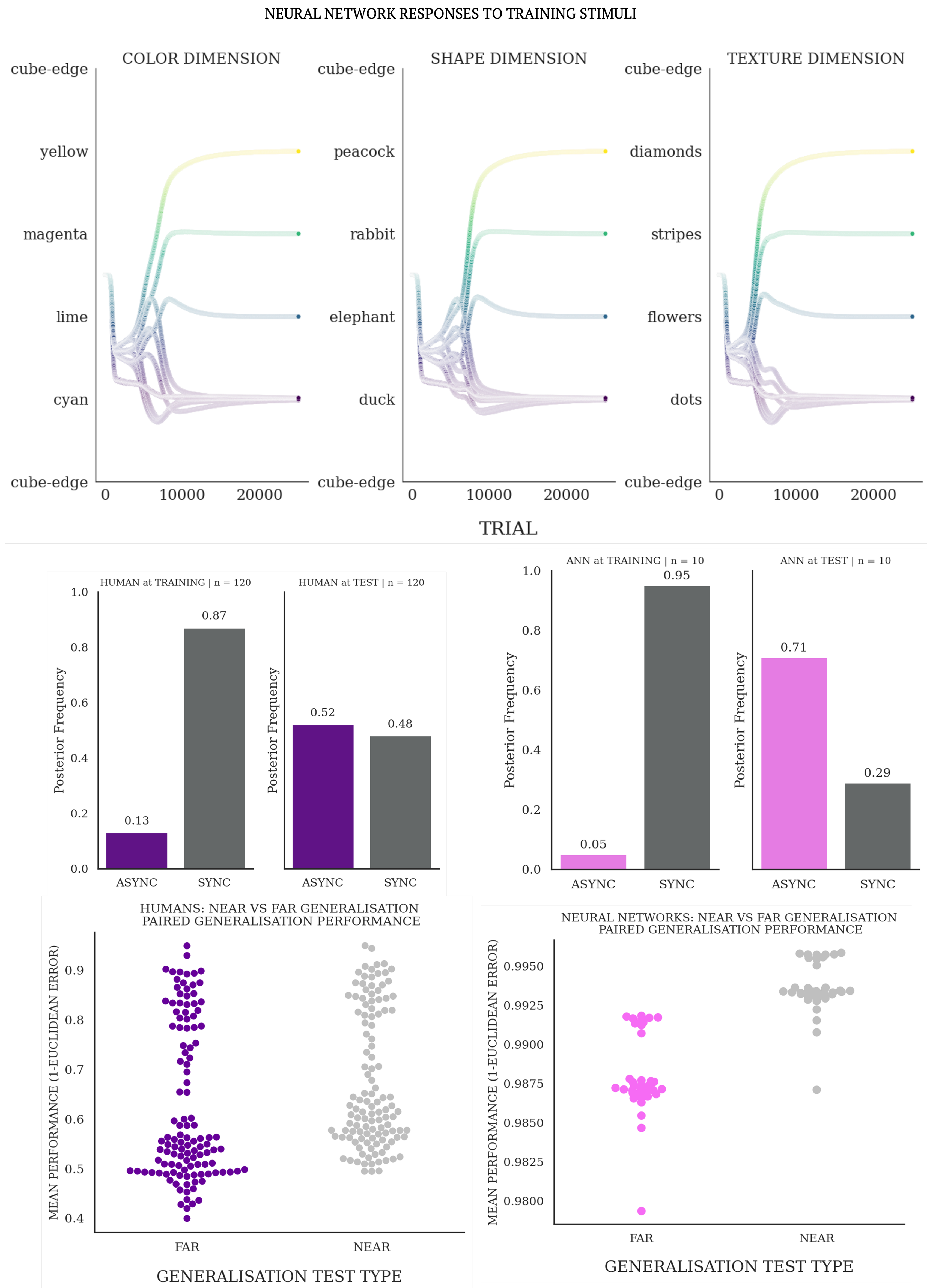

Figure 3. **Top**: Neural network's response locations to training stimuli over time for 3 dimensions: color, shape, texture, when nodal features are: cyan,duck,dots. In a first learning step predictions target the nodal stimulus and in a second learning step predictions develop a feature-specific distinction. Note asyncronicity across dimensions in the second step. **Middle**: Posterior frequency of asyncronous (AS) model fit to humans (left, purple) and neural networks (right, pink) in comparison to syncronous model (SY) at training and test. Both show asyncronicity at test.**Bottom:** average performance (inverted Euclidean error) across near vs far generalisation in humans (left) and neural networks (right). Both have higher performance for near generalisation than for far generalisation.

References

[Dekker et al., 2022] Dekker, R. B., Otto, F., and Summerfield, C. (2022). Curriculum learning for human compositional generalization. *Proceedings of the National Academy of Sciences*, 119(41).