Compositional Learning of a Numerical Reasoning Task in Artificial Neural Networks

Mia Whitefield, Sophie Arana, Christopher Summerfield
Department of Experimental Psychology, Oxford

Introduction

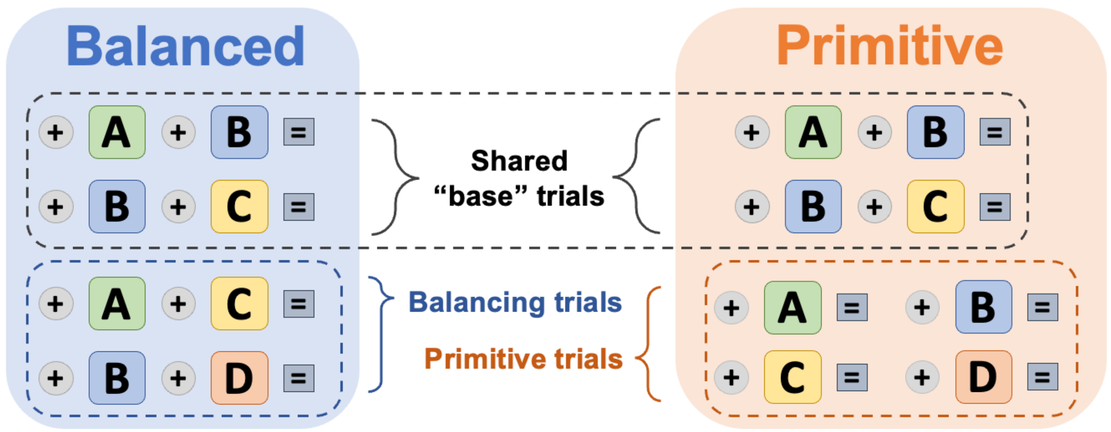
- **Composition**: the ability to recompose elements of prior knowledge to generate new mental operations or concepts.
- Humans excel at composition, usually after experiencing structured training (curricula), while current artificial neural networks struggle to match human performance.

Can training curricula encourage compositional learning in artificial neural networks?

Experimental Design

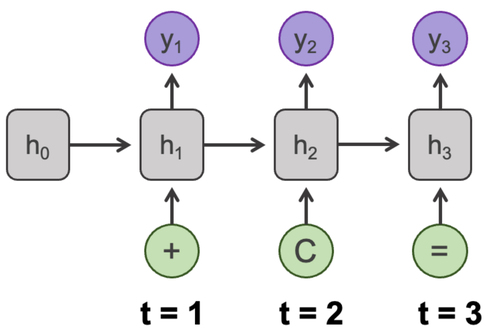
Symbolic arithmetic task

Curricula



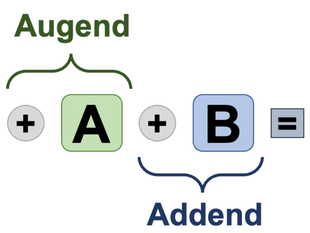
Models

- Recurrent neural networks (RNNs)
- Sequences of one-hot vectors represented sums
- One set trained on Balanced and one on Primitive curriculum

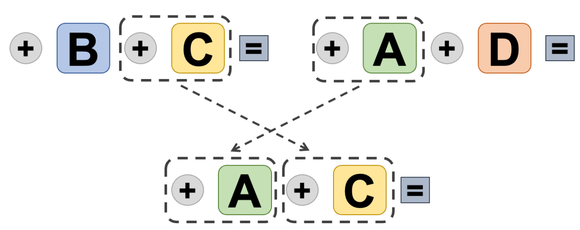


Experiment 1: Testing systematicity

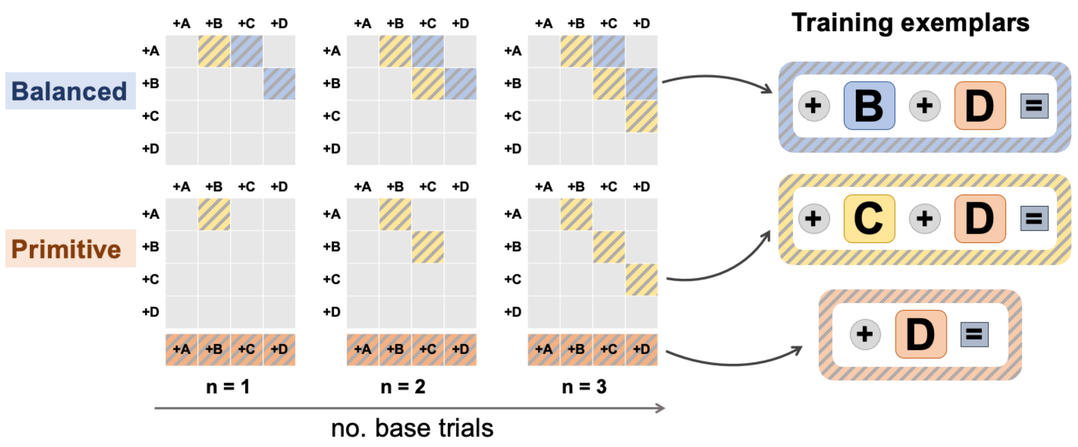
Simple addition



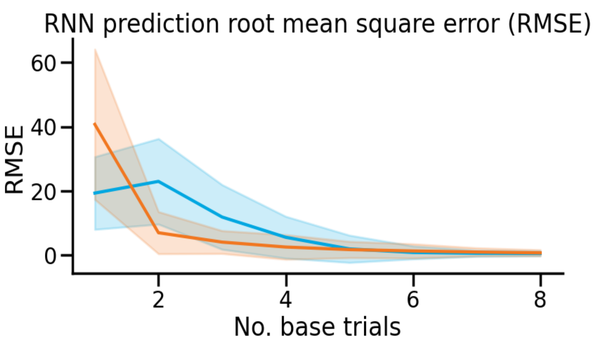
Systematicity

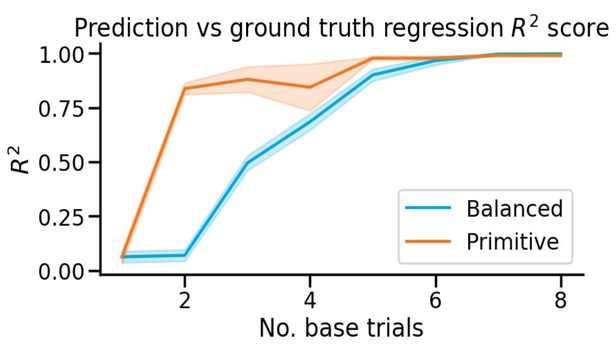


RNNs trained on increasing number of exemplars



- With few training exemplars, Primitive vs Balanced group generalised better to unseen trials, indicating greater sample efficiency



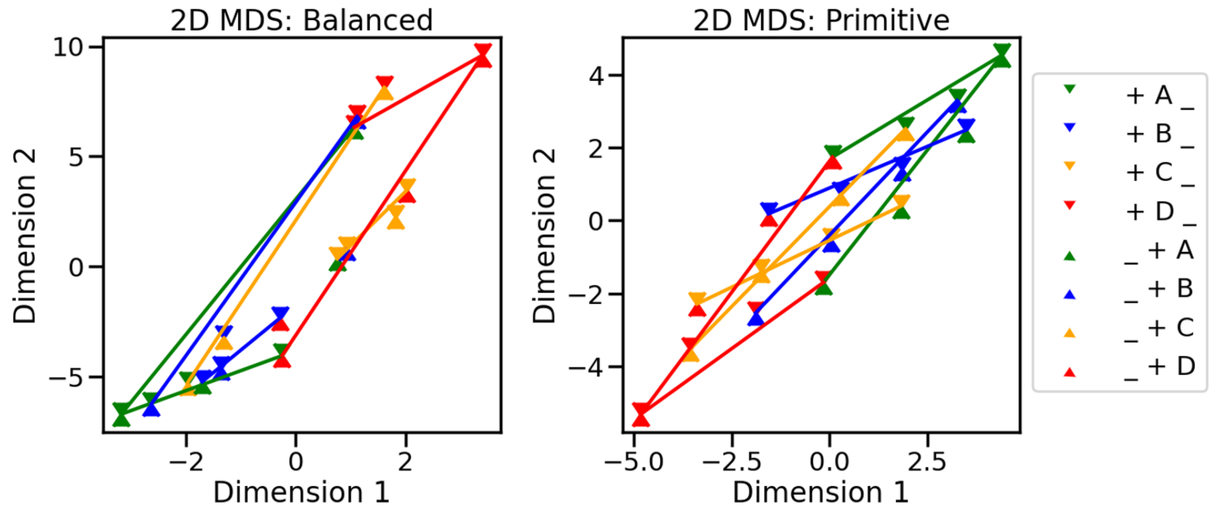


Representation Similarity Analysis (RSA)

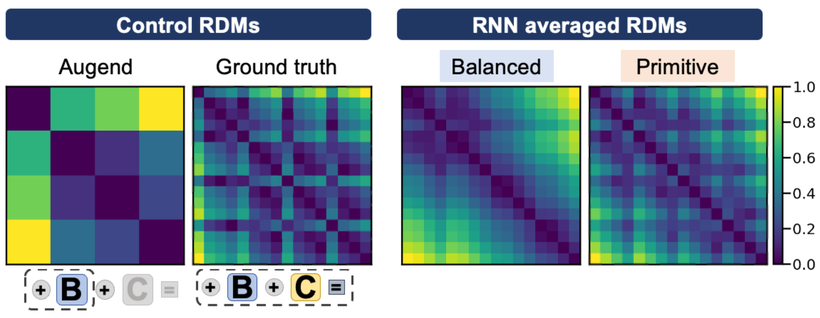
Multi Dimensional Scaling (MDS)

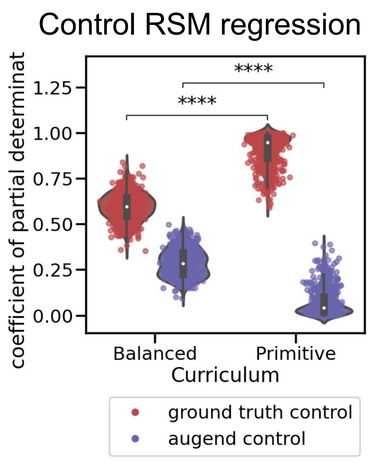
Hidden layer activations for the full set of 16 unique simple addition sequences

- Primitive group formed a sheared grid-like arrangement
- Balanced group clustered representations by the augend



Representation Dissimilarity Matrix (RDM) analysis

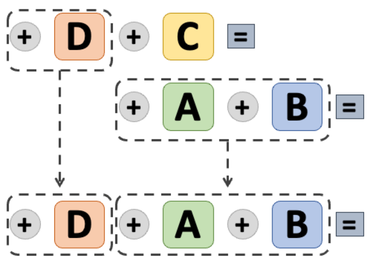


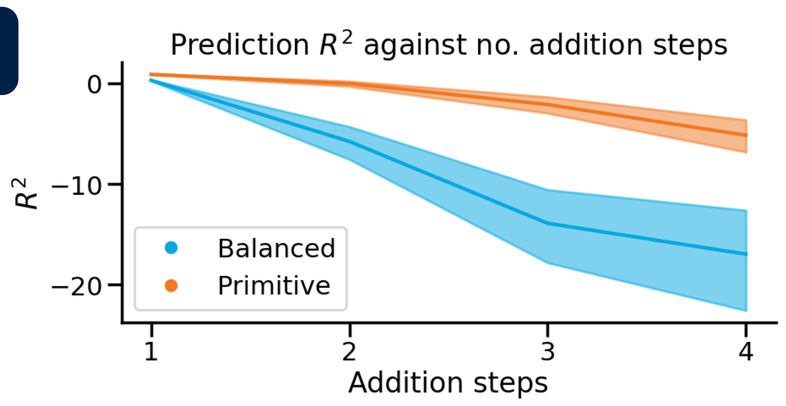


- Primitive group trial representations corresponded more closely to ground truth numerical differences than the Balanced

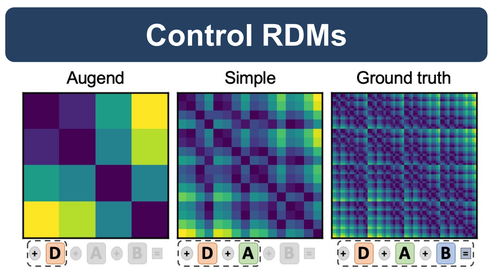
Experiment 2: Testing productivity

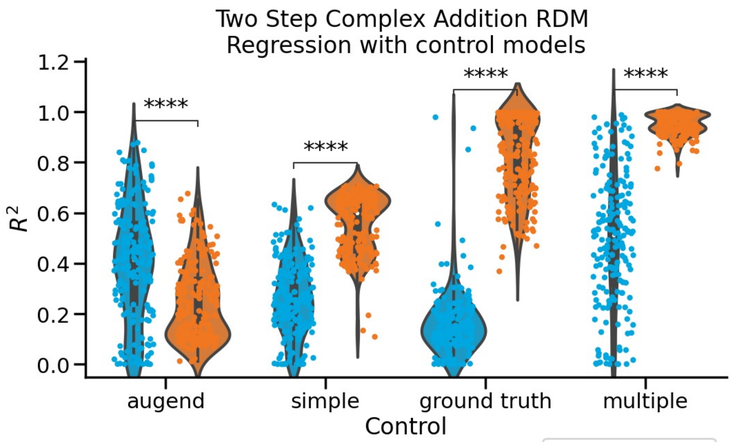
Productivity





- RNNs trained on simple addition trials (2 base trials) were evaluated on 2, 3, and 4 step complex addition trials
- Performance degraded with increasing addition steps for both groups, but drop was more pronounced for the Balanced.





- Primitive group RNNs generalised better to longer addition sequences

Conclusions

- Primitive training facilitated compositional generalization in neural networks in an arithmetic task.
  - improving sample efficiency, systematicity, and productivity.
- RNNs with primitive training vs without had distinct representational geometries that corresponded to the ground truth numerical differences more closely.