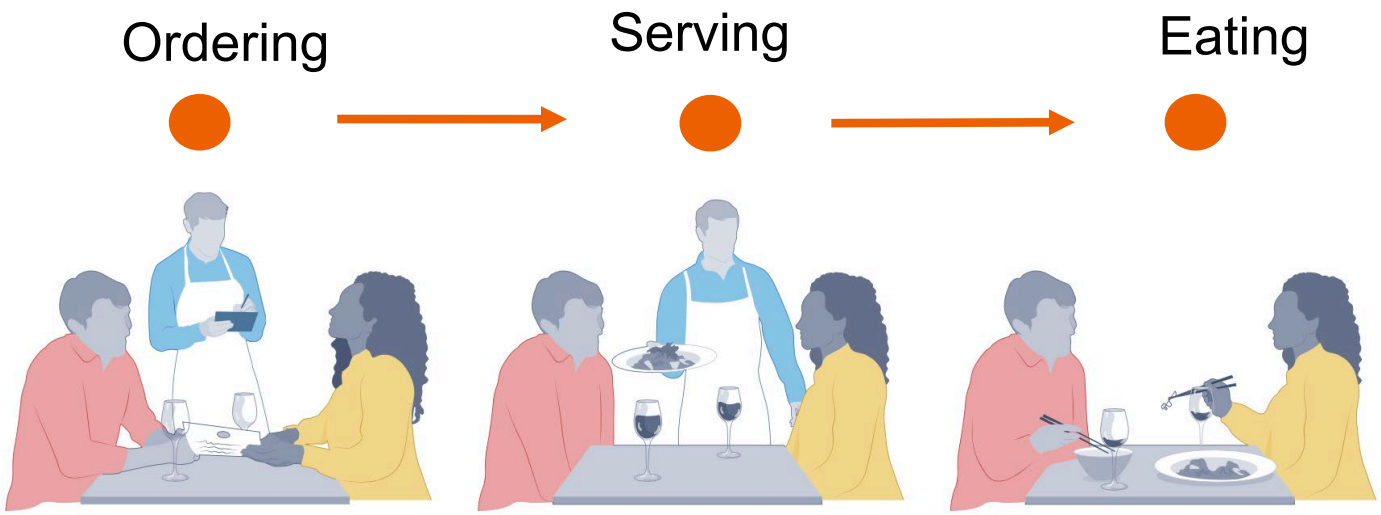
Shaping Sequence Attractor Schema in Recurrent Neural Networks

Zhikun Chu , Bo Hong ,
Xiaolong Zou*, Yuanyuan Mi*

NeurIPS 2025

Sequence schema

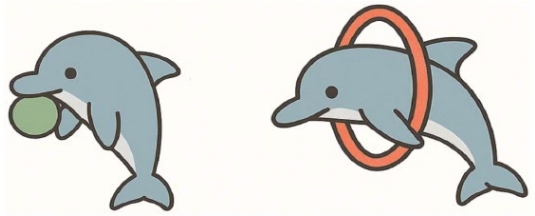
- A generalized and reusable abstraction of knowledge enabling flexible intelligence



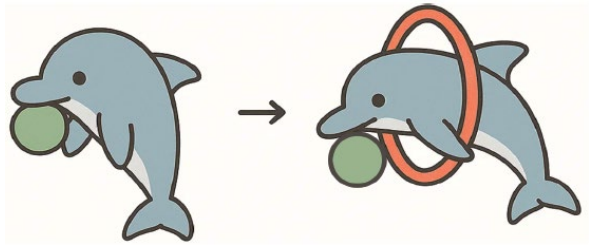
Generalizing the schema across different scenarios

How to learn a complex sequence schema ?

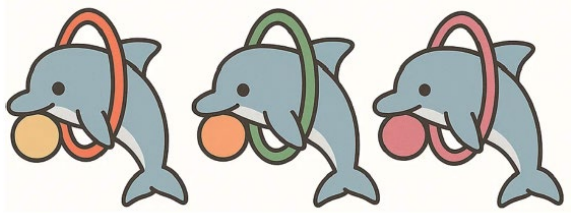
- Many tasks are too complex to be amenable to simple trial and error learning
- Shaping is critical: breaking complex sequence task into simple subtasks learned gradually



Task primitive learning



Task sequence learning



Task schema learning

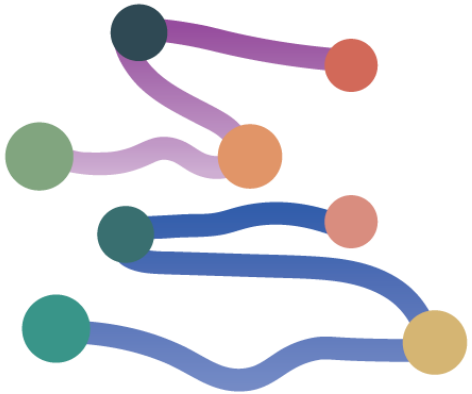
How sequence schemas are learned via shaping ?

Hypothesis: Sequence schemas are learned via shaping through sequence attractors. These attractor dynamics emerge gradually through the shaping process.

Task primitive learning:
Forming basic attractors for simple task parts

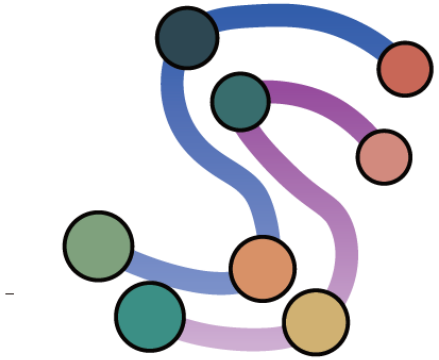
Discrete Attractors

Task sequence learning:
Linking individual attractors into sequence attractors



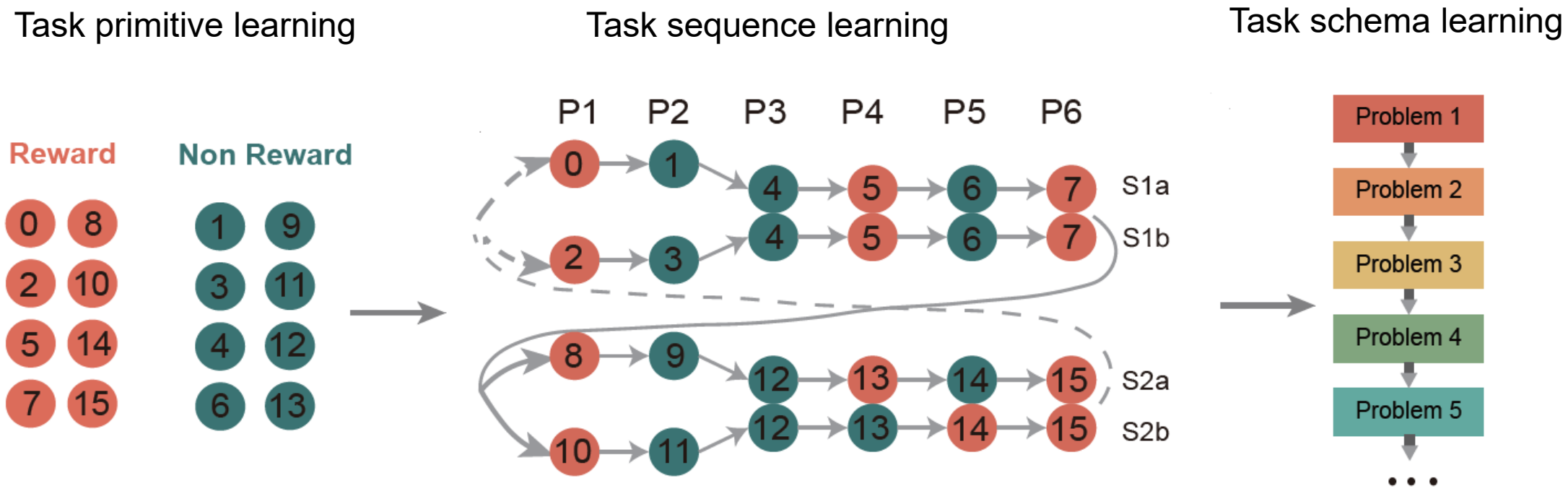
Sequential attractors

Task schema learning:
Compressing into a compact, abstract attractor structure



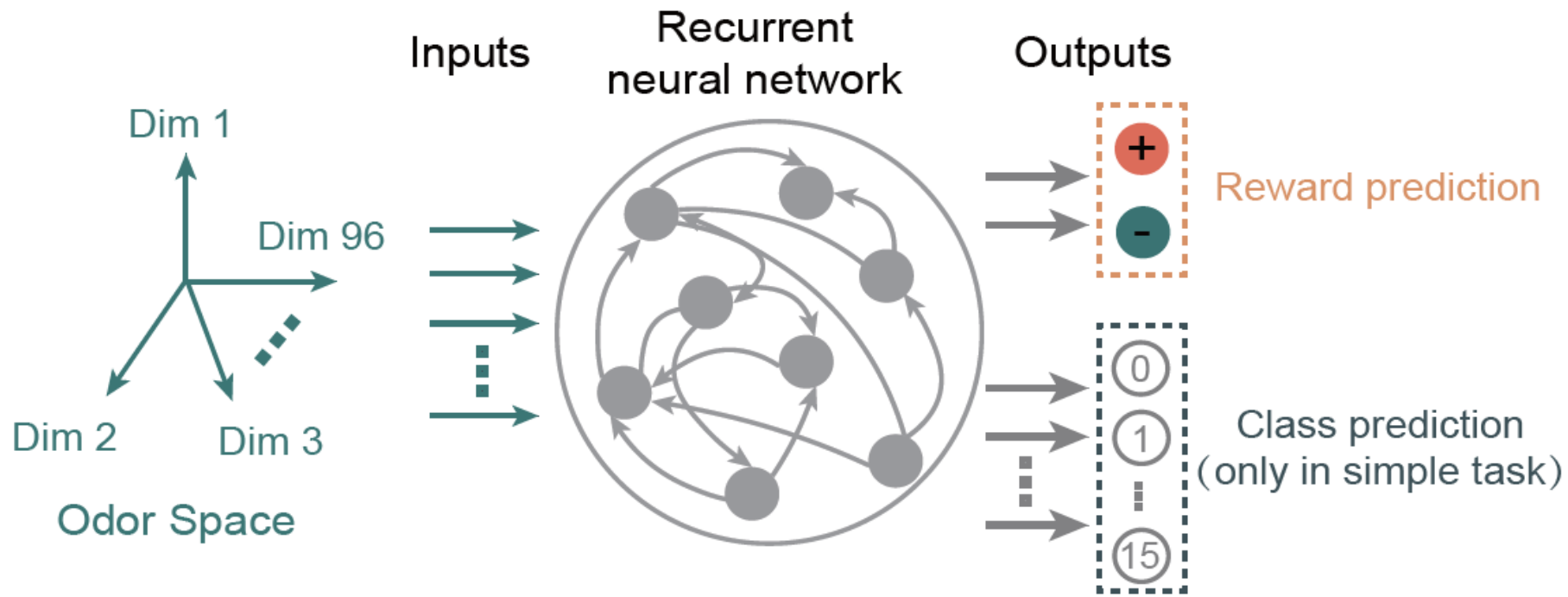
Attractor Structures

Shaping Paradigm in Animal Training



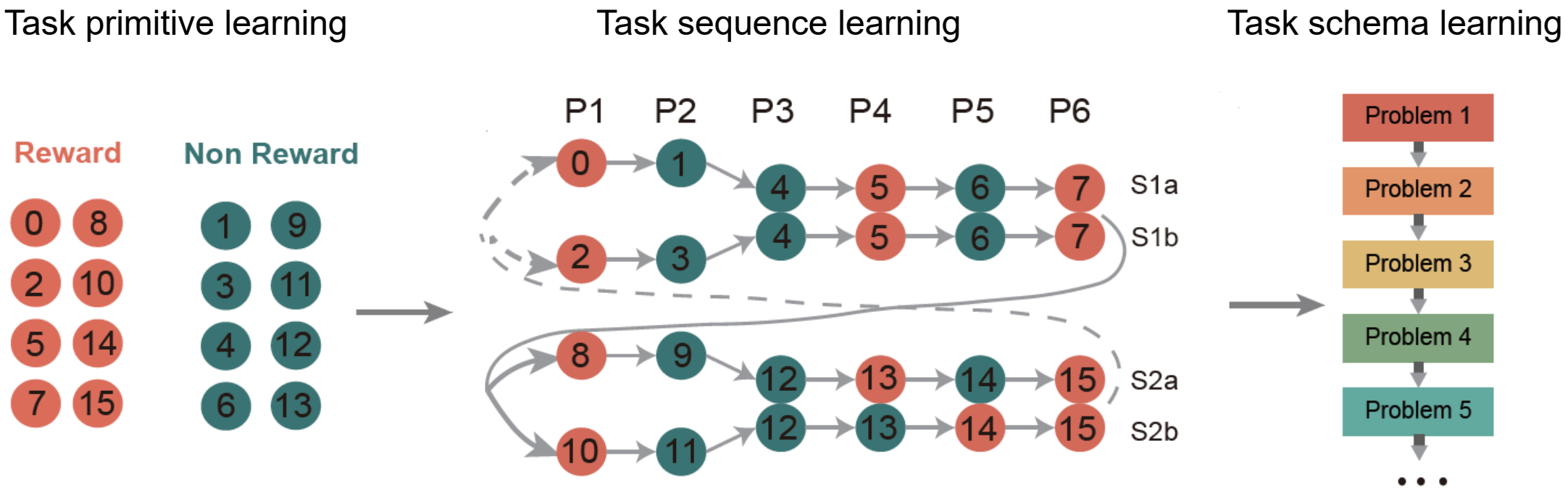
Zhou et al., *Nature*, 2021

Model architecture



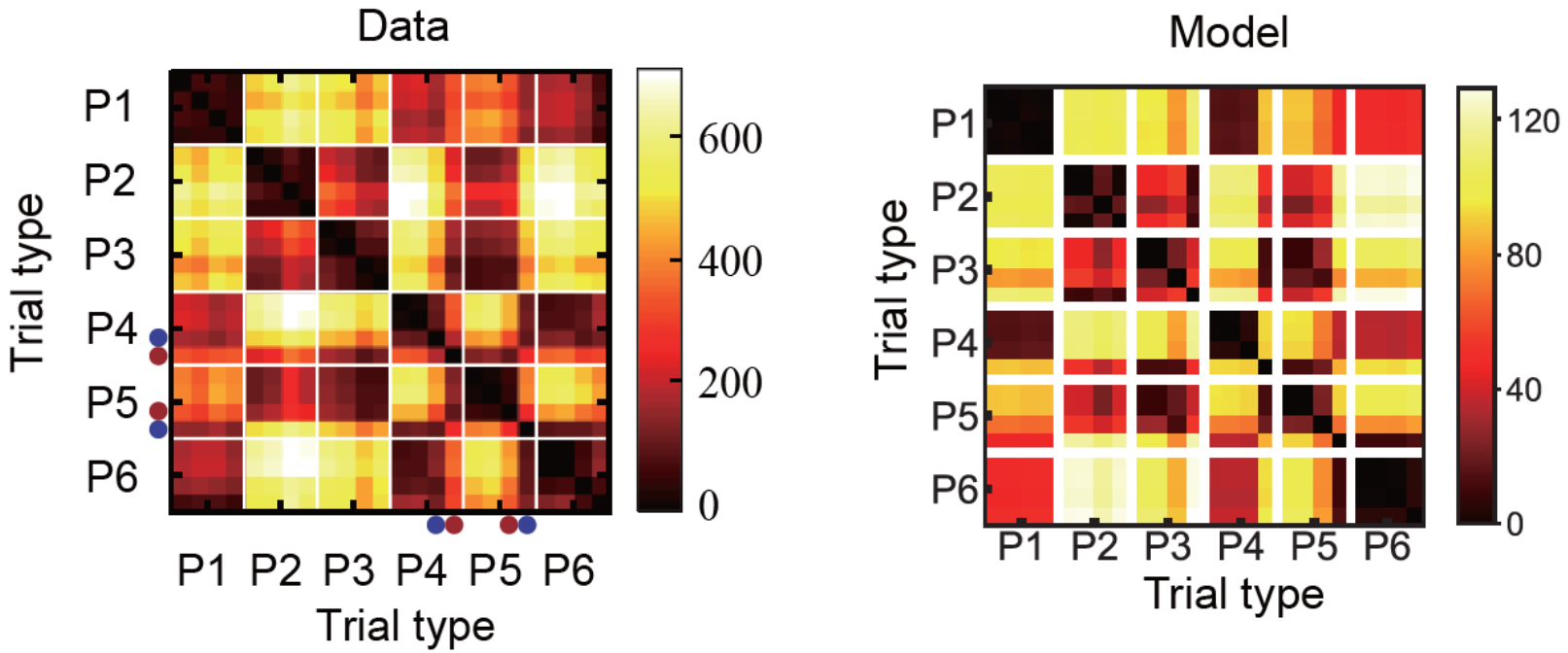
- Input : one-hot vector encoding of the current odour
- Target: predicts reward and/or classifies odours, depending on the training phase

Learning odor-sequence task via shaping



- Task Primitive Learning: RNNs learn to predict rewards based on individual odors.
- Task Sequence Learning: Pretrained RNNs are further trained on odor-sequence tasks.
- Task Schema Learning: The RNN is trained on multiple odor-sequence tasks sharing the same structure.

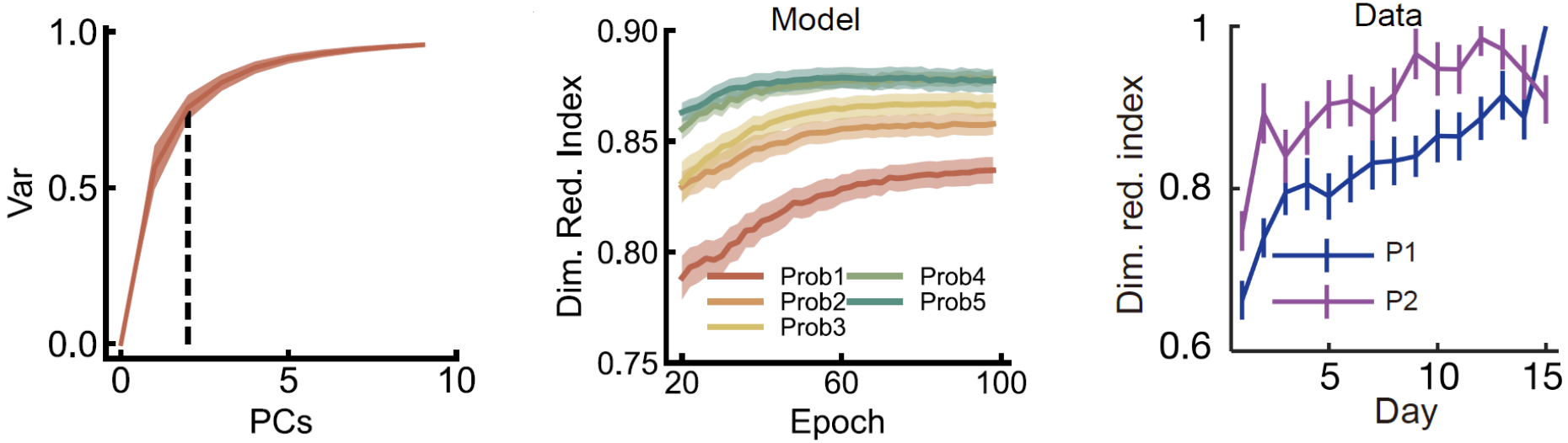
Reproducing Key Features of OFC Schema Learning : structured task representation geometry



- The RNN's representational dissimilarity matrix (RDM) closely aligns with that observed in rats' orbitofrontal cortex.

Zhou et al., *Nature*, 2021

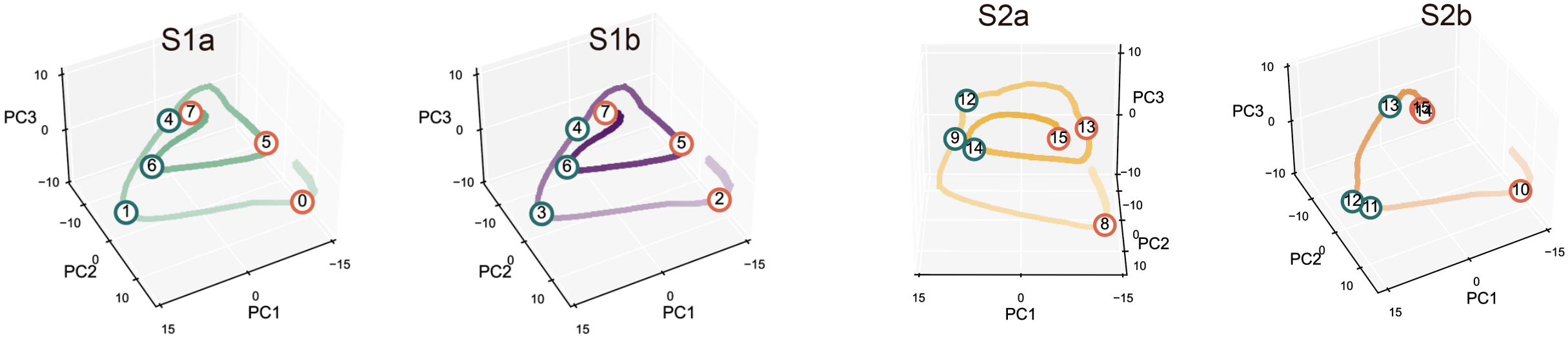
Reproducing key features of OFC schema learning: progressive compression of representational dimensionality



- As training progresses, the top 3 principal components (PCs) increasingly explain more variance in neural activity.

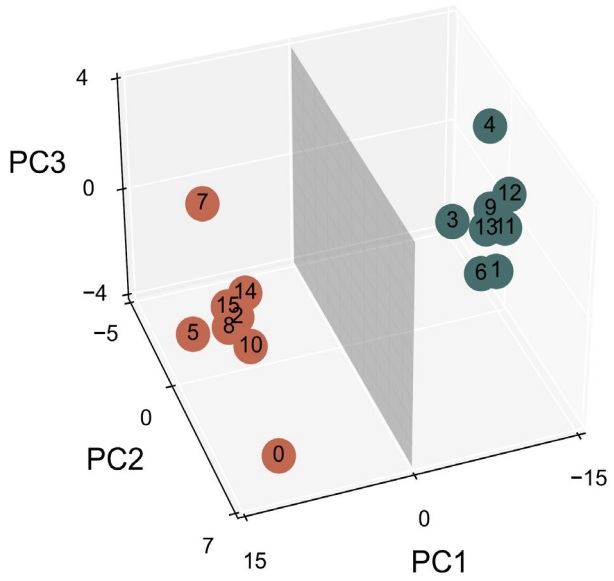
Zhou et al., *Nature*, 2021

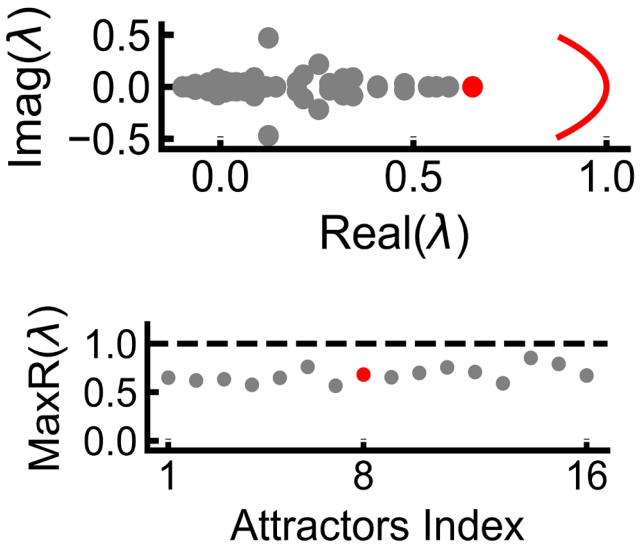
Four sequences form distinct trajectories with six fixed points each (one per odor position).



How sequence attractors are learned through shaping protocols?

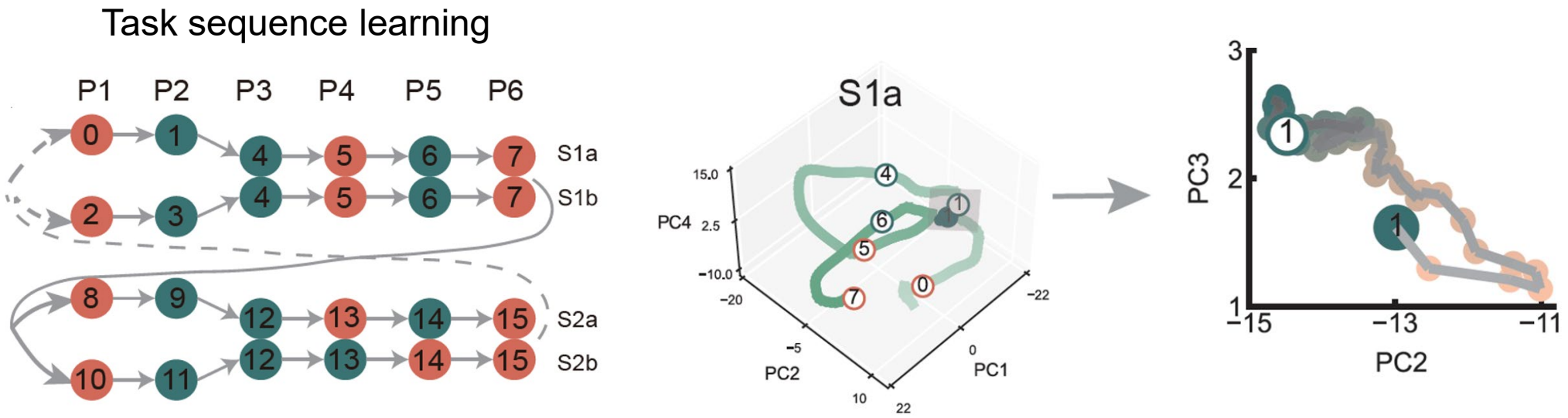
Task Primitive Learning — Formation of Discrete Attractors





- The network successfully maps odors to rewards.
- Network states form discrete attractors.
- Each odor corresponds to a stable network state.

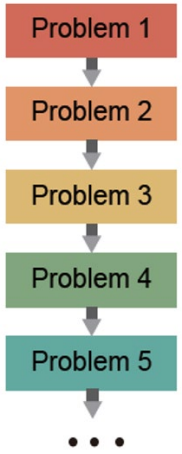
Task Sequence Learning — Linking discrete attractors to form sequence attractors



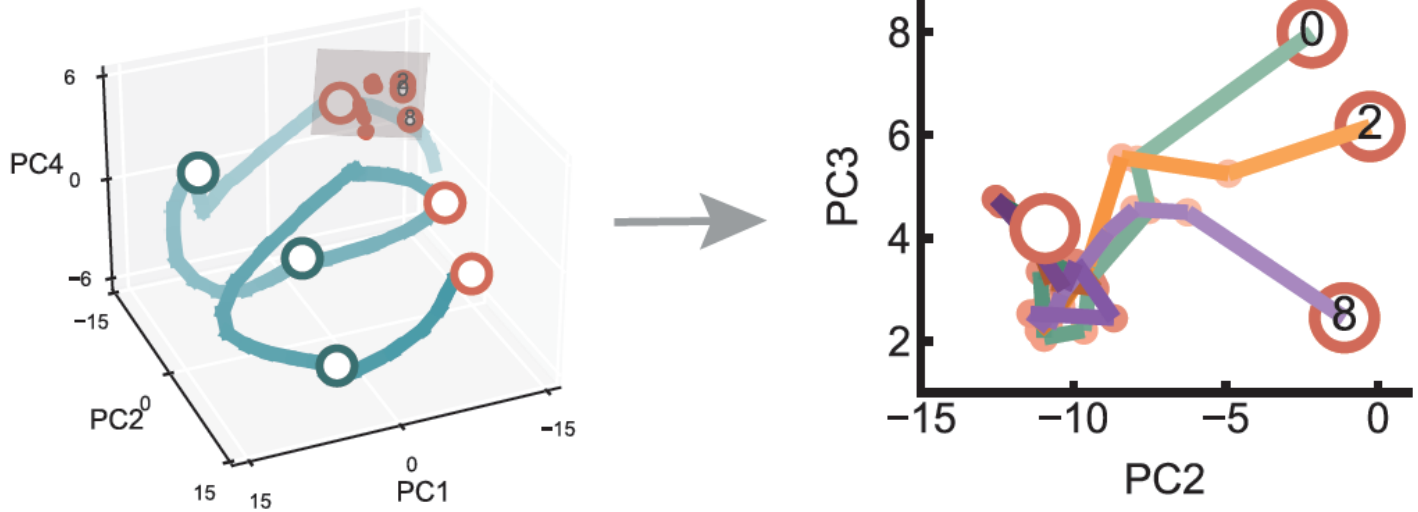
- Prior attractors from primitive learning migrate to form sequence attractors.
- This demonstrates attractors are reused, reorganized, and linked during sequence learning.

Task Schema Learning— Abstraction and Generalization

Task schema learning

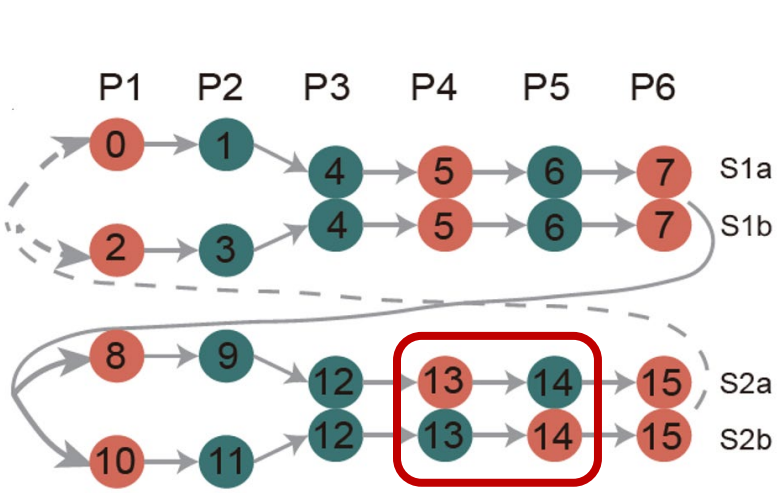


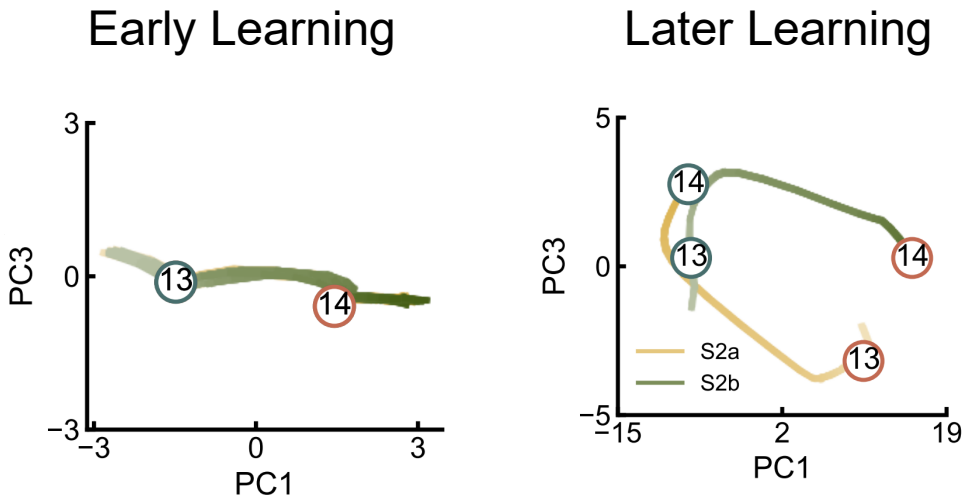
shared attractor trajectory



- Trained on tasks with shared structure but different odors
- Network representations shift from odor-specific to abstract, low-dimensional manifolds

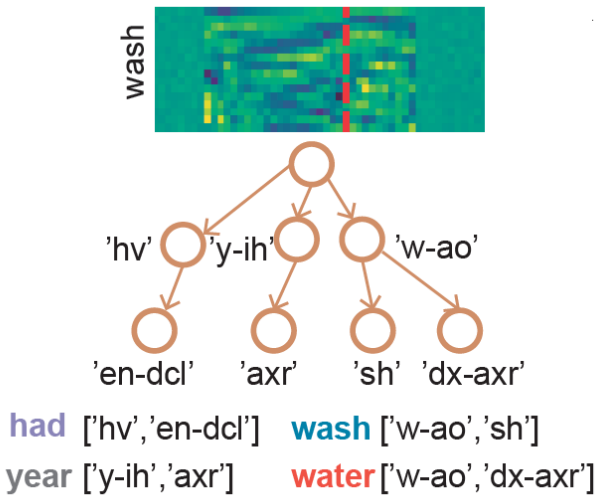
Shaped and unshaped RNNs follow divergent learning trajectories

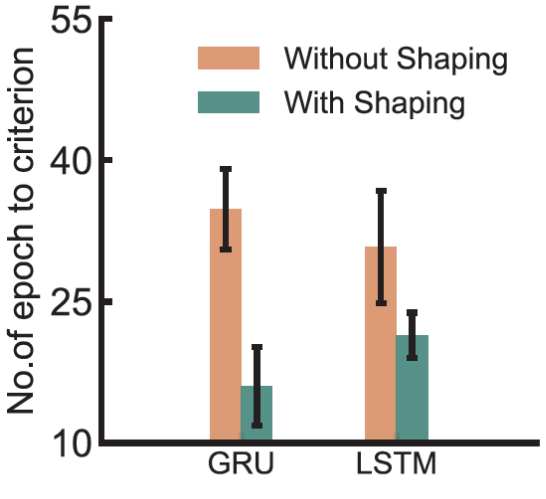


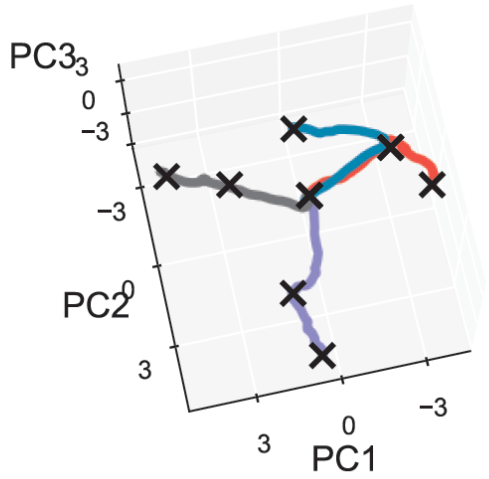


- In RNNs lacking shaping, the S2a and S2b attractor trajectories first collapse into one before separating, which contrasts with shaped RNNs.

Sequence Attractor-Based Shaping Improves Learning Efficiency in Keyword Spotting







- Applied our shaping framework to spoken-word recognition.
- Three-stage training: phoneme primitives → word sequences → abstract schema.
- Shaping enables faster convergence and interpretable attractor structure.

Conclusion:

- We systematically replicate key behavioral and neural features of schema learning observed in the rat OFC.
- We show that sequence schemas can be encoded as **sequence attractors**.
- We identify a novel dynamic process of schema formation: from **point attractors → sequence attractors → abstract sequence attractors**.