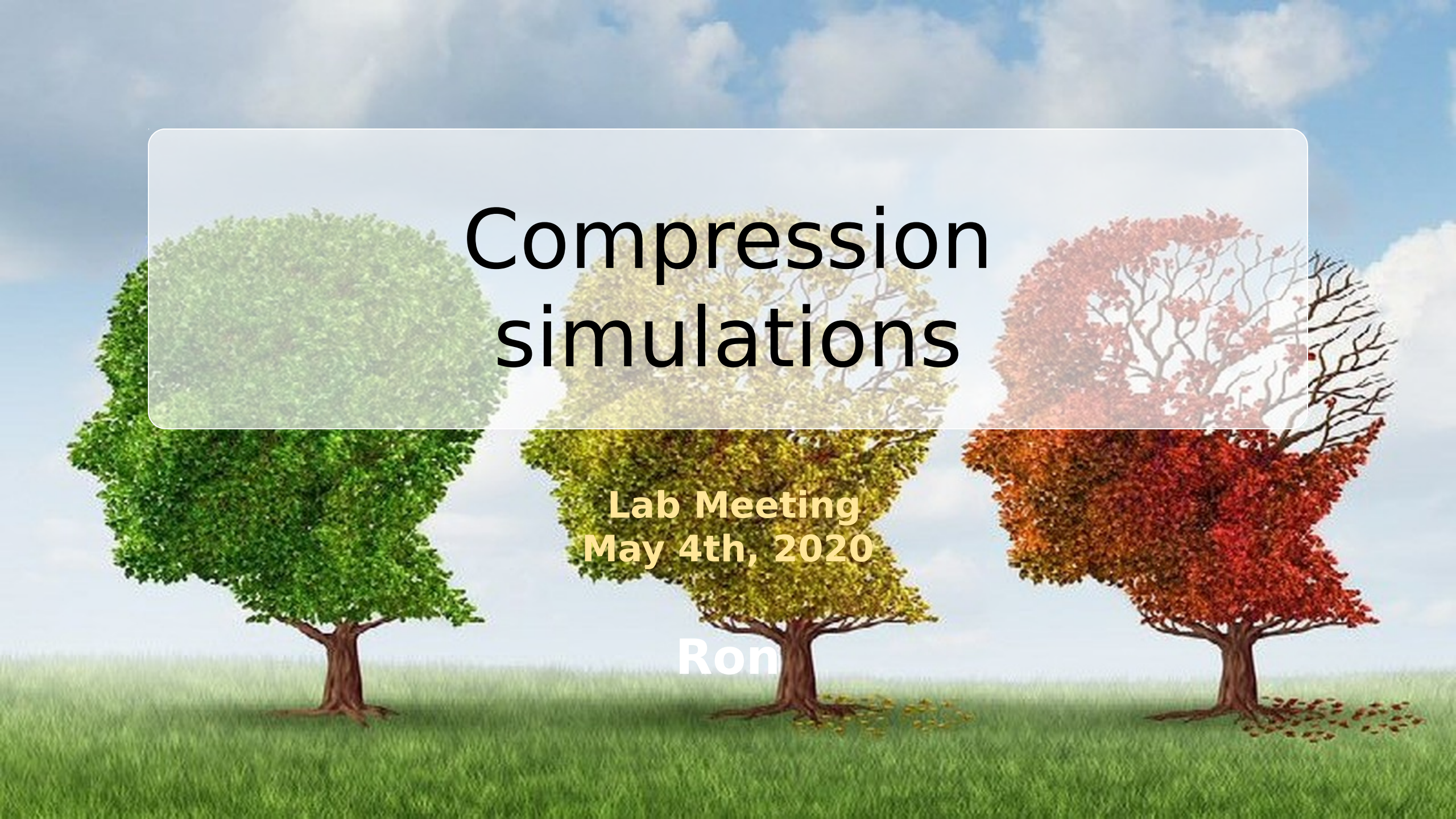


Compression simulations

Lab Meeting
May 4th, 2020

Ron

What is this?

Answer: a blue banana

That was easy

Humans have little difficulty in interpreting novel scenes.

Don’t need experience with blue bananas, or elephants with cat-like fur, to be able to recognize or imagine them.

Why?

non-useful representation

Property	Value
Pixel [0,0]	(255,255,255)
Pixel [0,1]	(255,255,252)
...	...

useful representation

Property	Value
Shape	Banana
Color	Blue
...	...

Why that was easy

Human perception does not consist exclusively of end-to-end optimization of explicit task-relevant supervision labels

- Human perception/learning:
  - Involves compressed representations (dimensionality reduction)
  - Involves compositionality (understanding “blue beach ball” and “yellow banana” -> “blue banana” becomes meaningful)
    - Potential contributor to superior continual and transfer learning in humans
    - Likely built up over a wealth of experiences in diverse settings which share some structure

Key questions

- How to model human-like compositional cognition?
- What are the benefits?
  - Few-shot learning of new categorization rules?
  - Extrapolation to unobserved regions of parameter space at test?
  - In general, does understanding the latent generative structure of the world provide inductive biases which explain at least some of the things humans can do, but machines currently cannot?

Agenda

- Compression simulations
- Weather-prediction task (behavior)

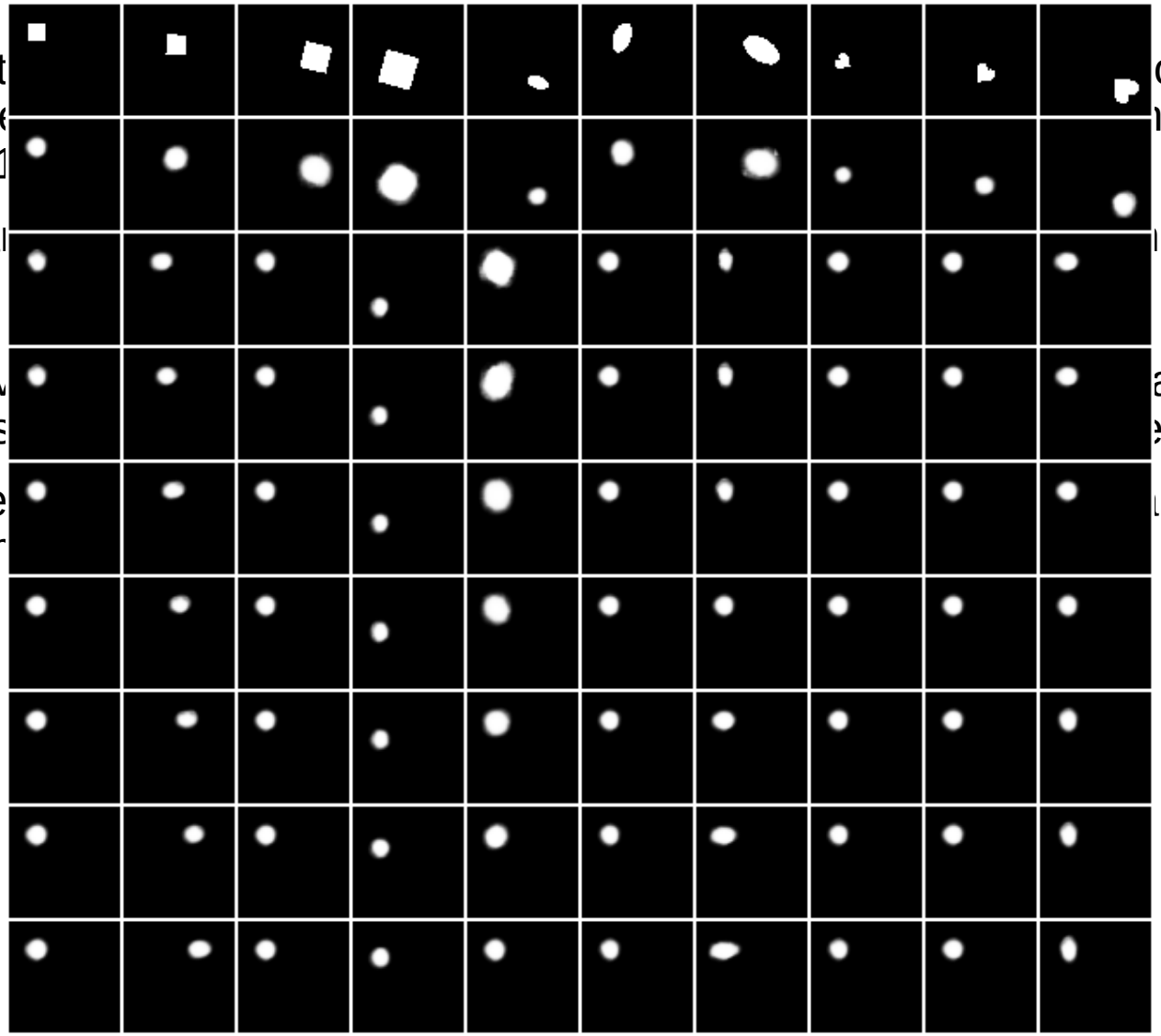
Compression/cont.learning setup

Human cognition likely employs a set general primitives that can be engaged and combined flexibly to rapidly learn new tasks

- How do we model this in a neural network?

1. With a simple linear neural network, train on different tasks.
2. After each individual task, store *n* ‘task modules’.
3. Then, evaluate on an integration task that combines the individual tasks in some way. Here, the compressed model only has to learn a few weights for how much each of its modules are used. In other words, learn composite task as linear combination of of stored task modules.

Related work



(1) In t of cells) is modified by similarity betwee nded locations and features on the other 1

Simila factors for these features

(2) ß-V angled representations that have been s eral cases 2

Furthe nsfer learning across tasks (upcor

1 Martinez-Trujillo (2005)
2 Higgins et al. (2017)
3 Achille et al. (2018)

Complementary learning systems

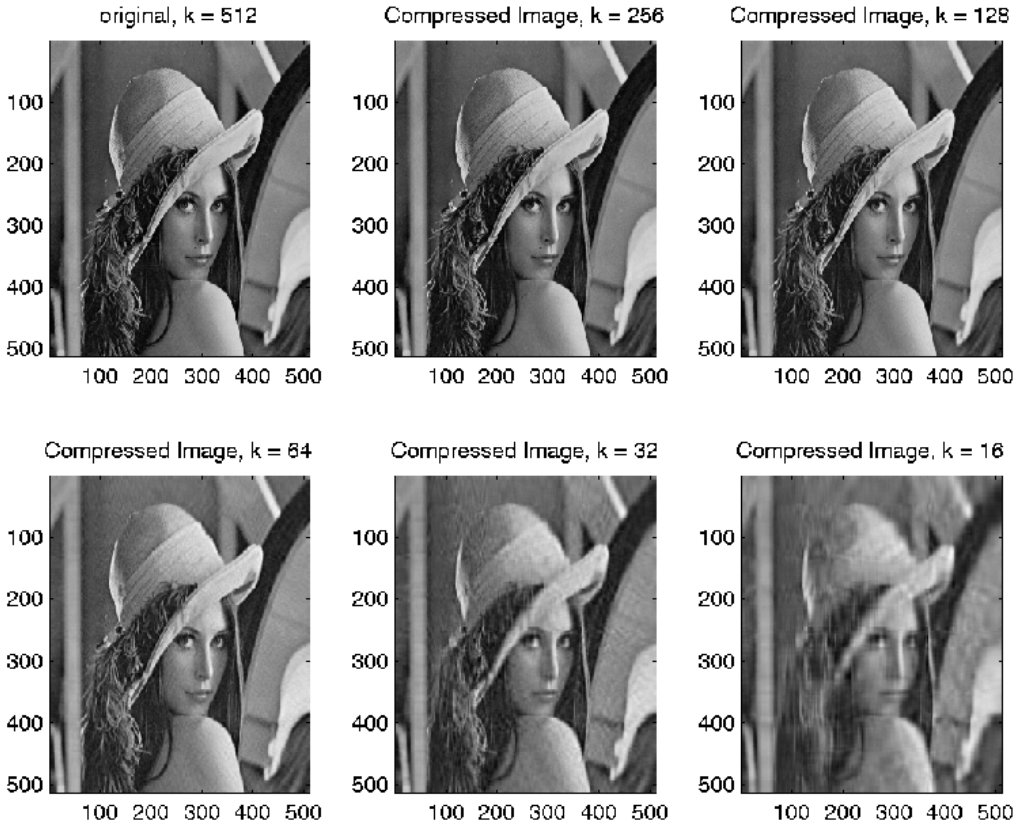
- Complementary learning systems (CLS) theory1:
  - Damage to HCA disrupts recent, but not remote memories
  - Hippocampus: fast, pattern-separated learning
  - Neocortex: integrates over episodes. Slow learning & overlapping representations

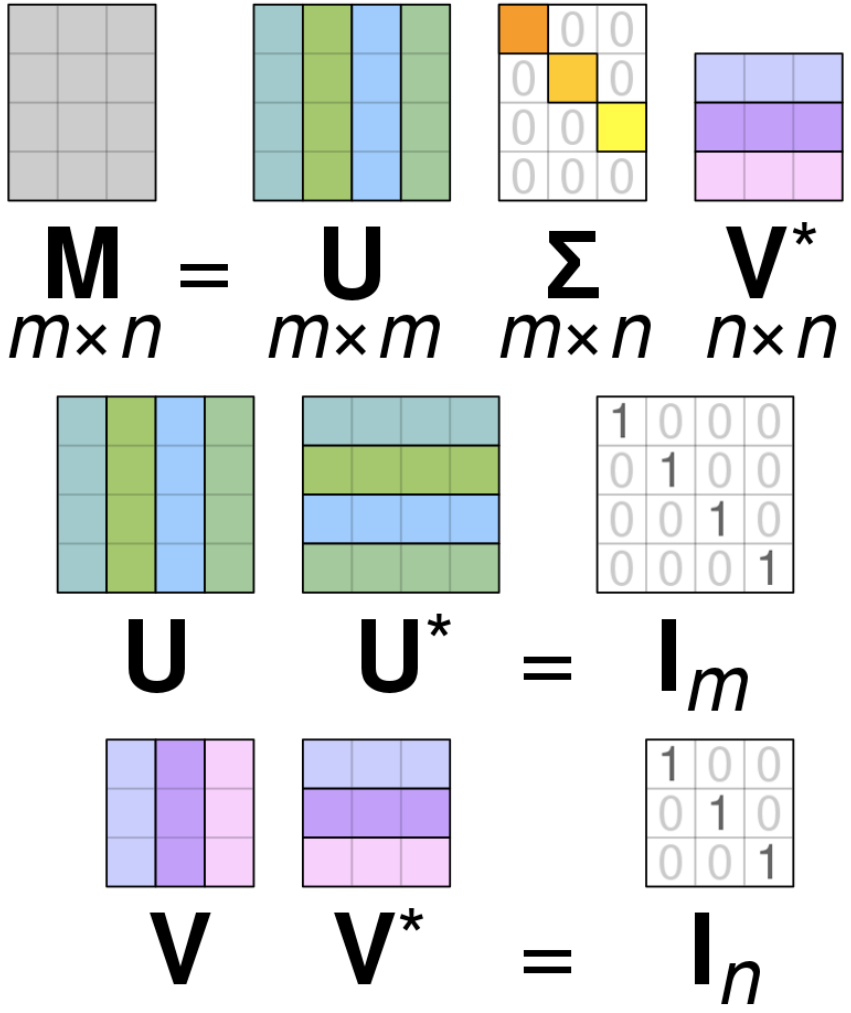
- Neural networks:
  - Single learning system
  - New experiences overwrite old ones (catastrophic interference)
  - Transfer is limited:
    - Generalization to novel inputs
    - Limited ‘learning-to-learn’ of related tasks

1 McClelland, McNaughton, and O'Reilly (1995).

Singular vector decomposition(SVD)

- SVD: matrix factorization method: M=UΣV
- Singular values in Σ ordered by magnitude: compressed by taking k<min(m,n) σs





Implementation

- Decomposing a task into an orthogonal basis set hopefully gives rise to generalizable primitives

General algorithm:

- Learn elementary task using linear neural network (‘hippocampal learning’)
- Store the first *k* components of an SVD compression of the matrix product of the layer weights, then reset weights
- We could then express new tasks using only k*n weights (n the number of tasks and k the number of SVD components per task) (‘neocortical learning’)
- With starting weights 1, initial behavior is meaningful and represents an ‘average policy’

Test case 1: MNIST

- Elementary tasks: 1-vs-all classification.

Ex. task 1: “is it a 0 or not?”

Task 2: “is it a 6 or not?” etc.

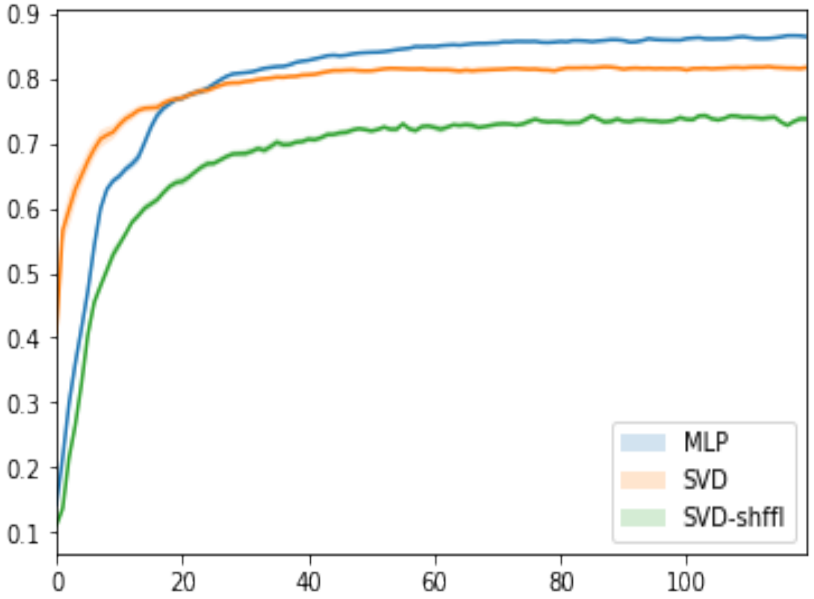
- Composite task: 10-way classification

“Which digit is it?” (0-9)

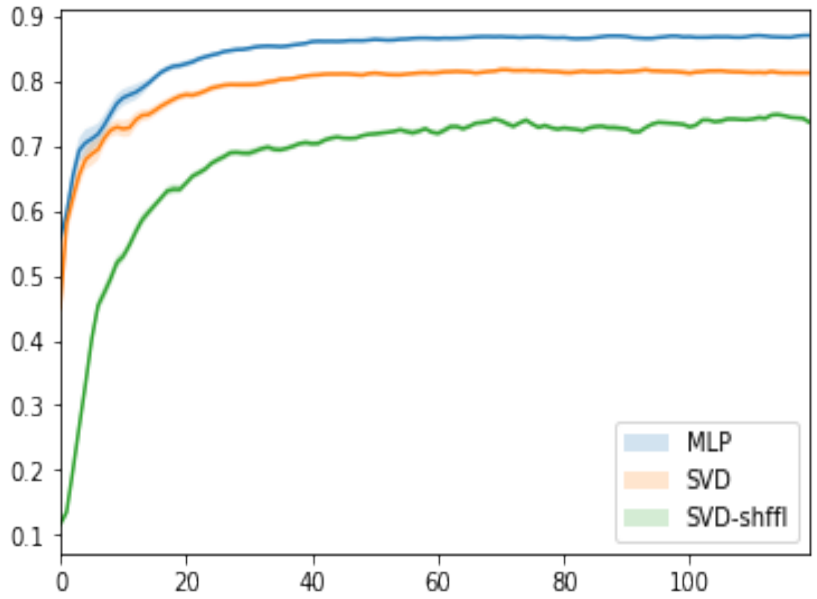
Test case 1: MNIST

- Individual tasks: 1-vs-all classification. Composite task: 10-way classification

MLP initialized at composite task

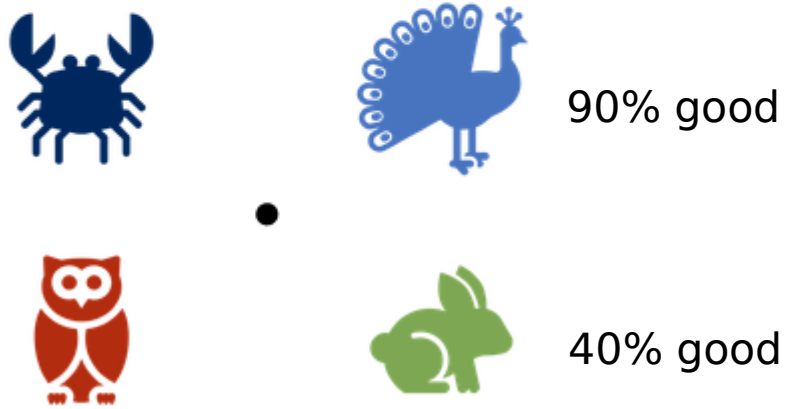


MLP trained on individual tasks in sequence



Test case 2: weather prediction task

- Original by Yang & Shadlen (2007)
  - Monkey LIP firing rates correspond to adding up per-stimulus log-likelihoods
- Curriculum condition: train on 1- or 2-stimulus displays
- Parallel condition: train on 4-stimulus displays
- Test: always 4-stimulus displays
  - Some combinations are left-out at training



==> good

Questions

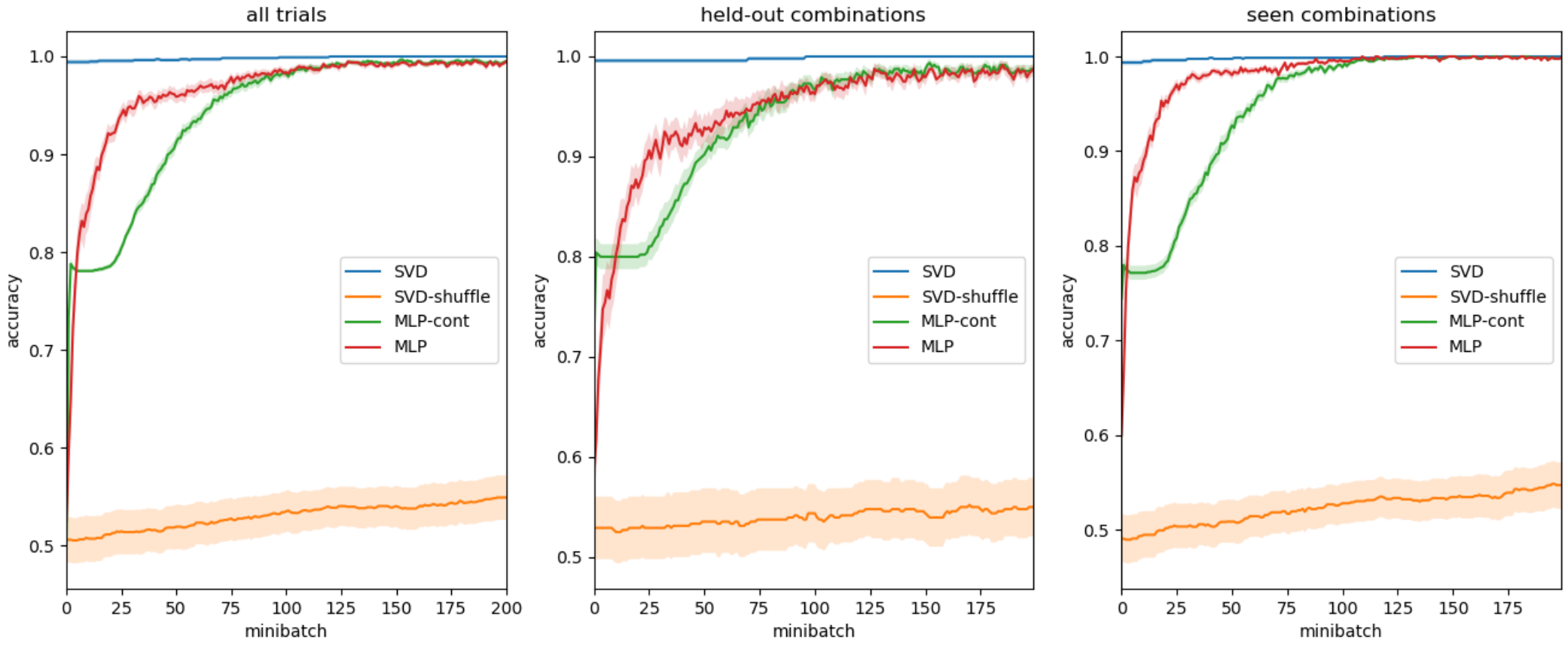
- Individual tasks: One stimulus (present or absent) or two stimuli per task
- Composite task: 4-stimulus arrays

Questions:

- Is there a benefit to compression in terms of learning speed for the integration task?

- The compression network should be able to generalize to any input combination. How do the compression and vanilla networks compare on held-out combinations?

Results



- Initial behavior is sensible integration over past tasks (conceptually similar to passive dynamics LMDP).
- Mild generalization advantage to held-out combinations.

Interim conclusions

- Is there a benefit to compression in terms of learning speed for the integration task?

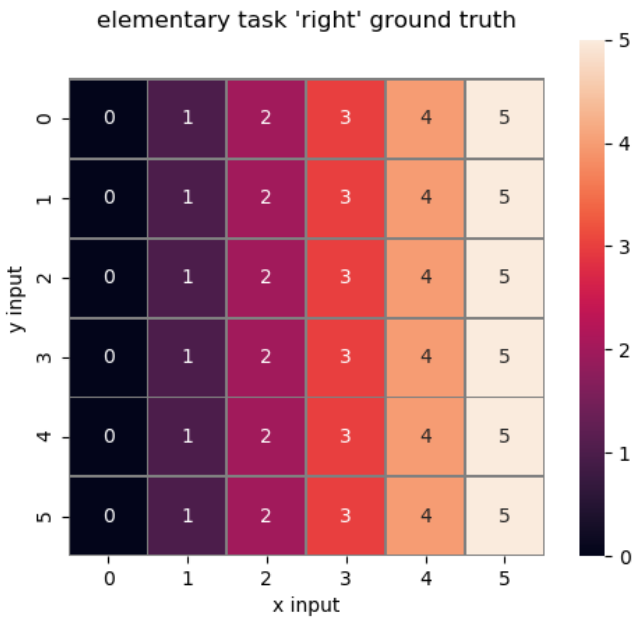
Yes, the SVD network performs virtually optimally even with its initial behavior

- The compression network ‘carves nature at its joints’ and should be able to generalize to any input combination. How do the compression and vanilla networks compare on held-out combinations?

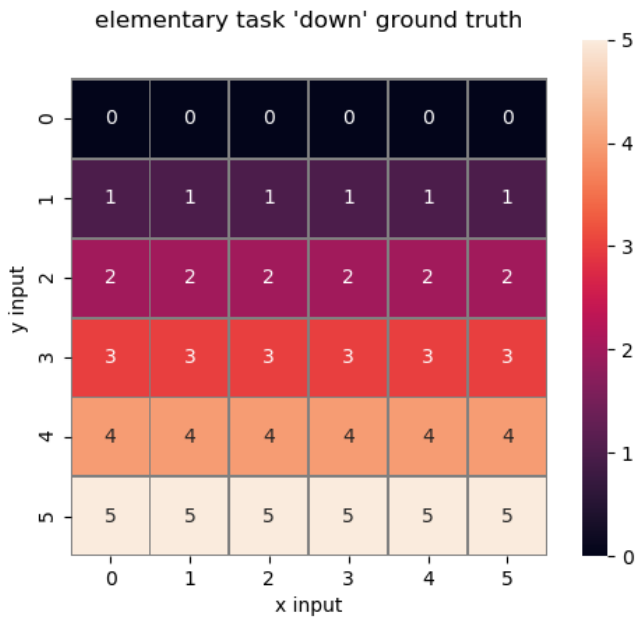
Yes, the compression network does better at out-of-sample generalization (modest differences in this setting)

Test case 3: Extrapolation – setup

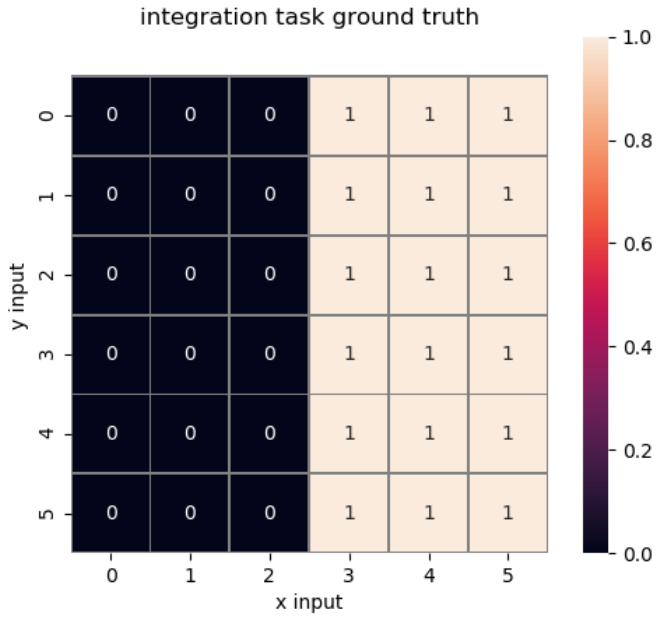
Elementary task 1



Elementary task 2



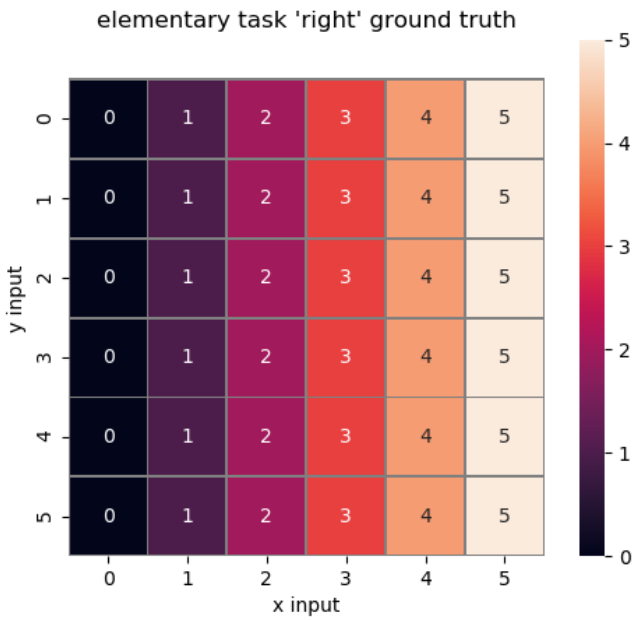
Integration task



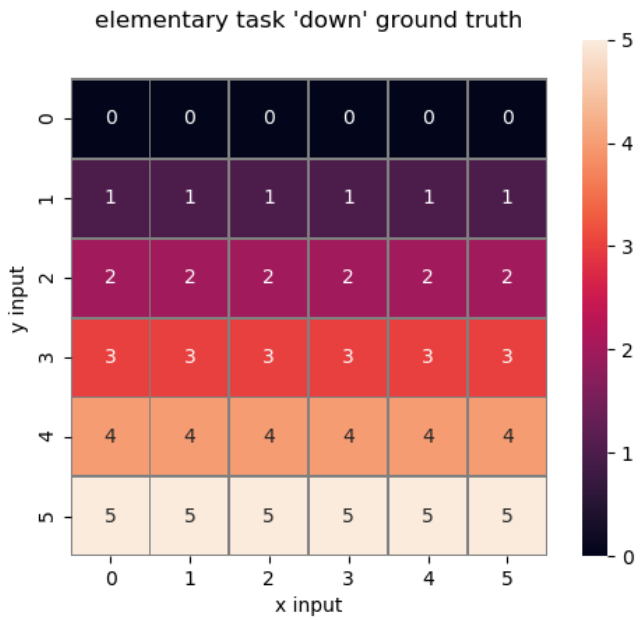
→ The integration task is a unidimensional categorization task based on one of the extracted features

Limited # test exemplars

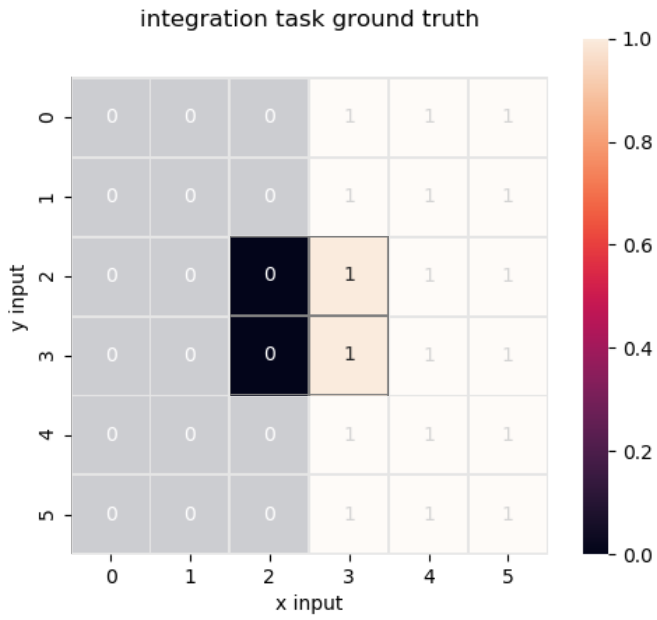
Elementary task 1



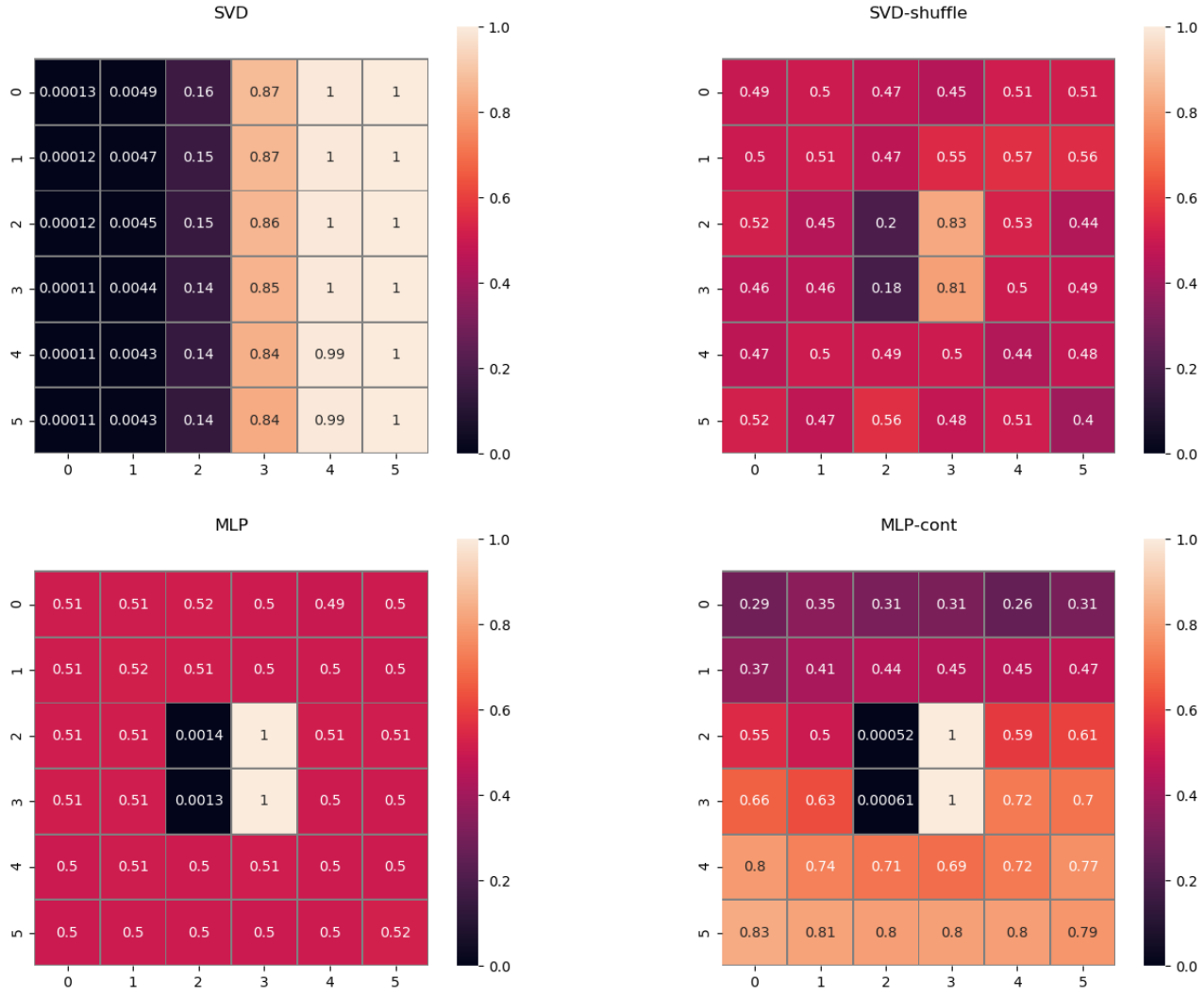
Elementary task 2



Integration task

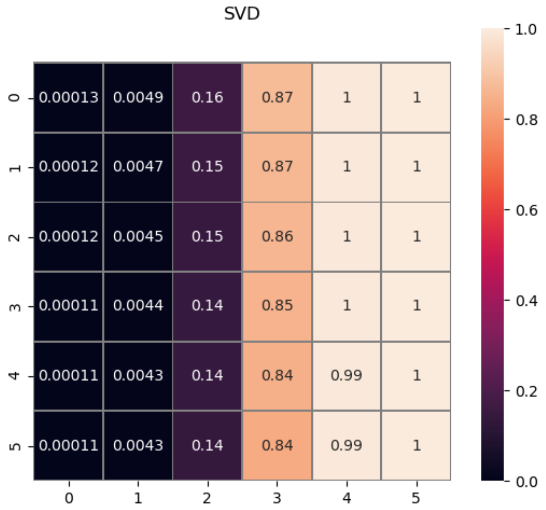


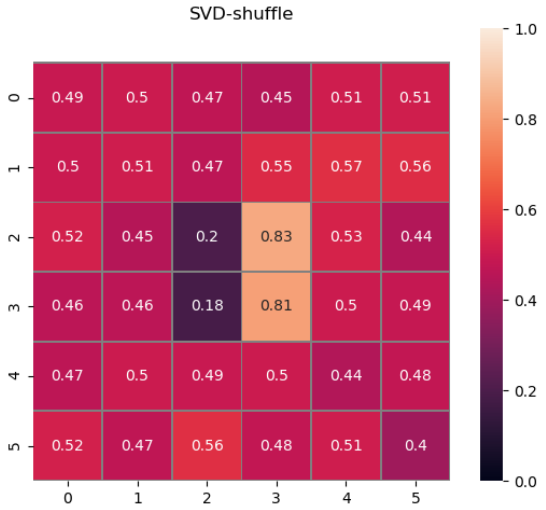
Qualitative test behavior



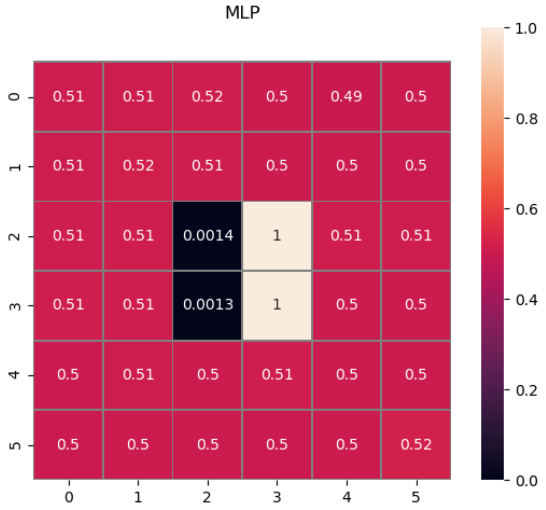
Qualitative test behavior

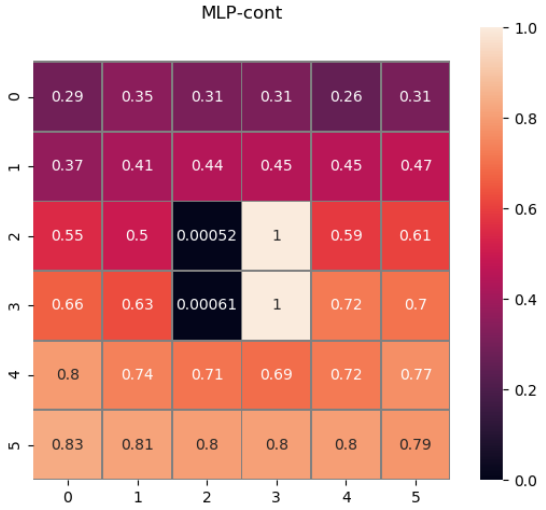
Only one with good extrapolation perf.





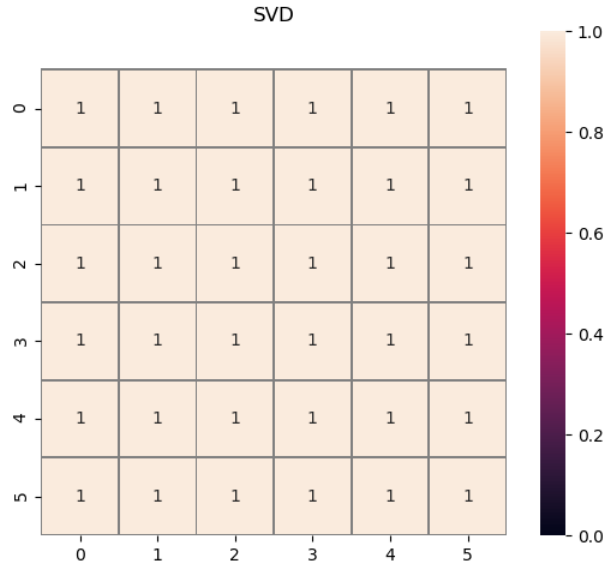
At chance for out-of -distribution ex.

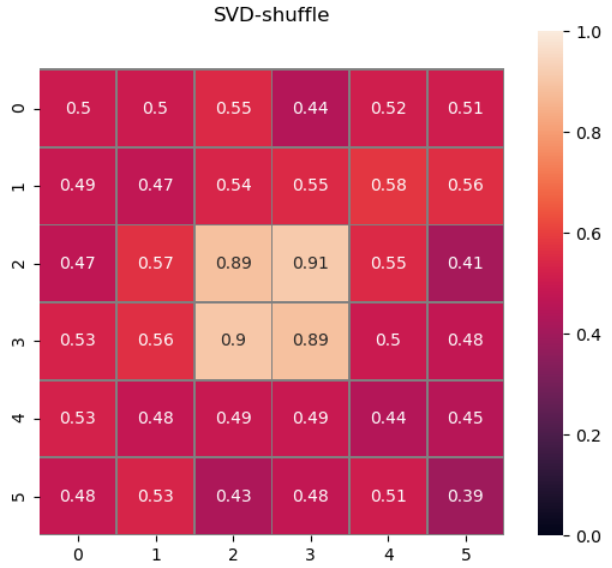


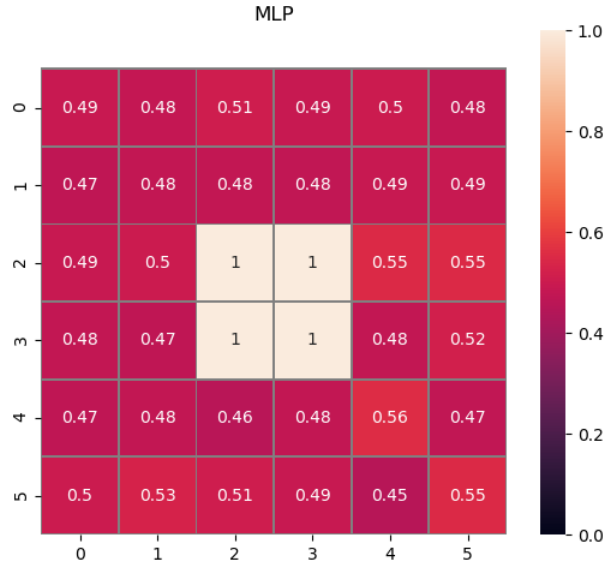


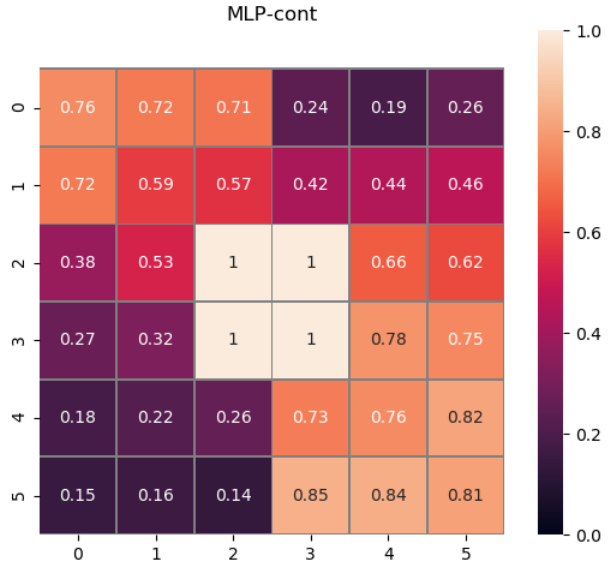
Overfits to most recent task

Confusion patterns



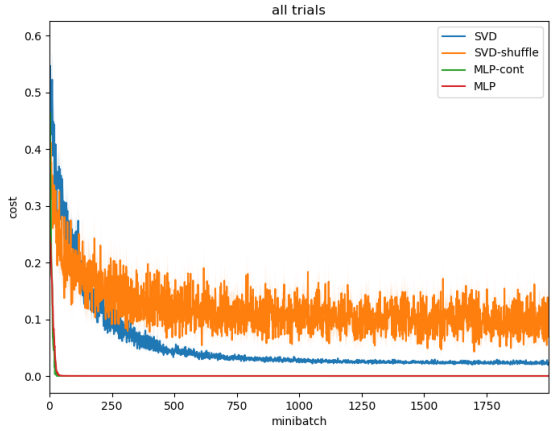




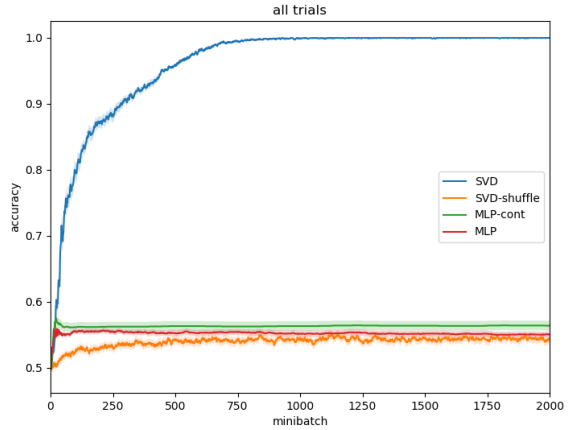


Temporal dynamics

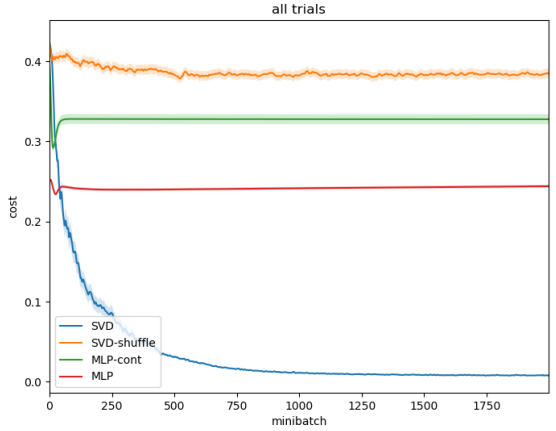
Integration training loss



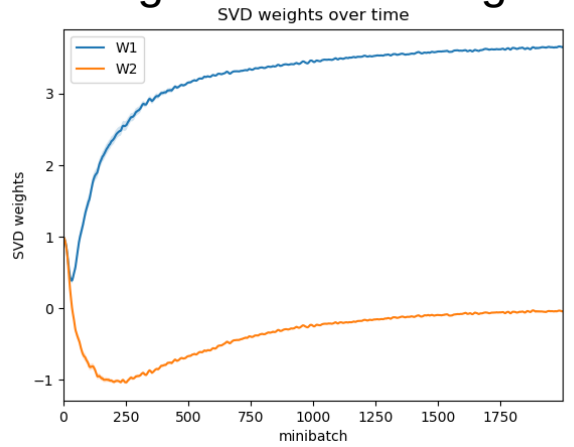
Integration test accuracy



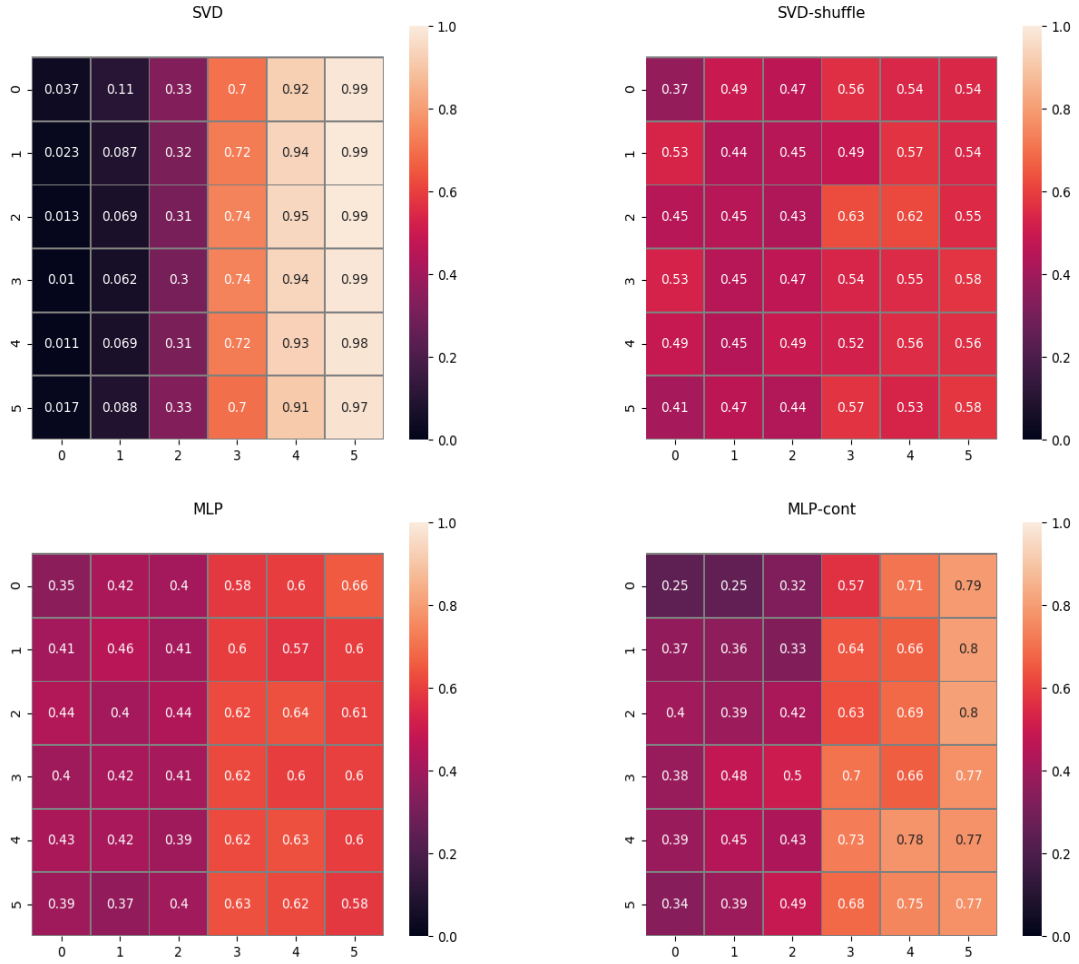
Integration test loss



Integration SVD weights



Control 1: random exemplars



Is handpicking exemplars that straddle the boundary essential? → use *k* random exemplars instead (k=8 in the figure)

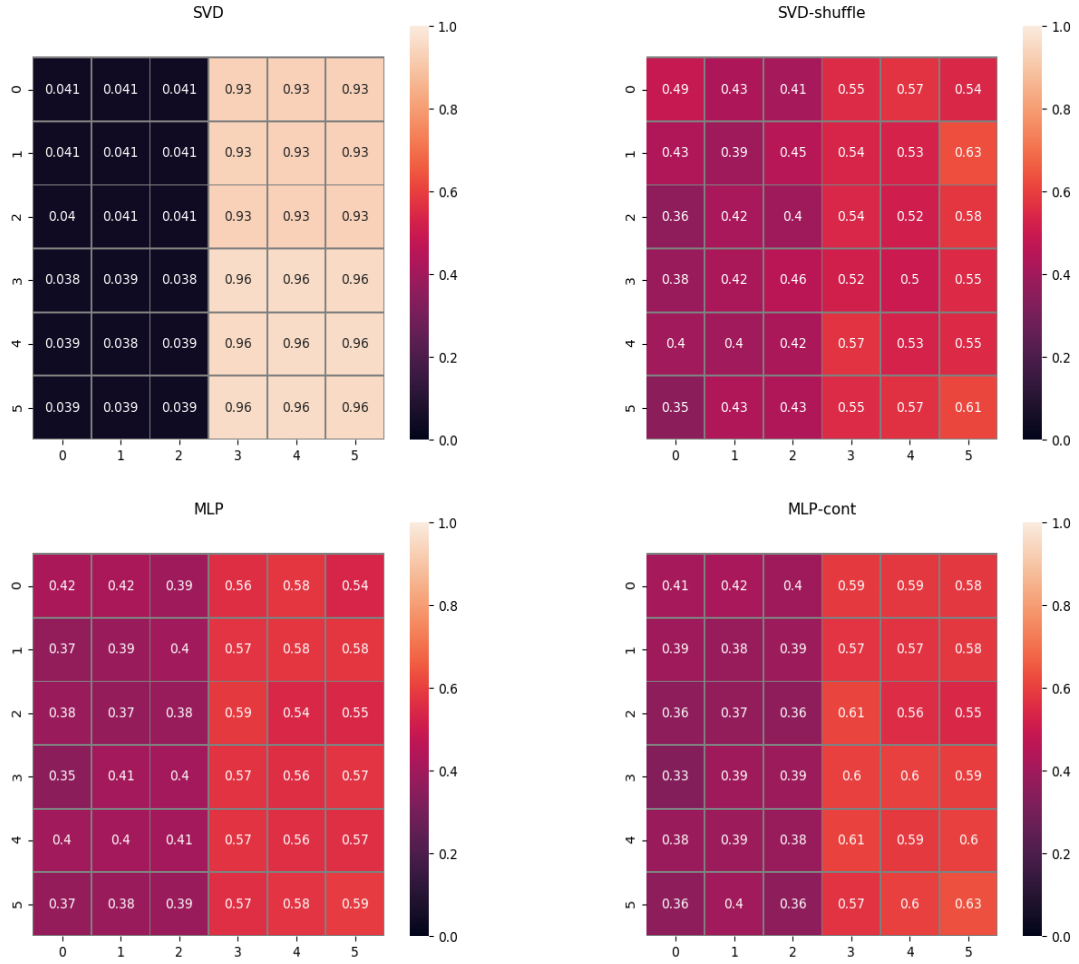
Furthermore: randomize elementary task order

→ Same results hold in this control setting

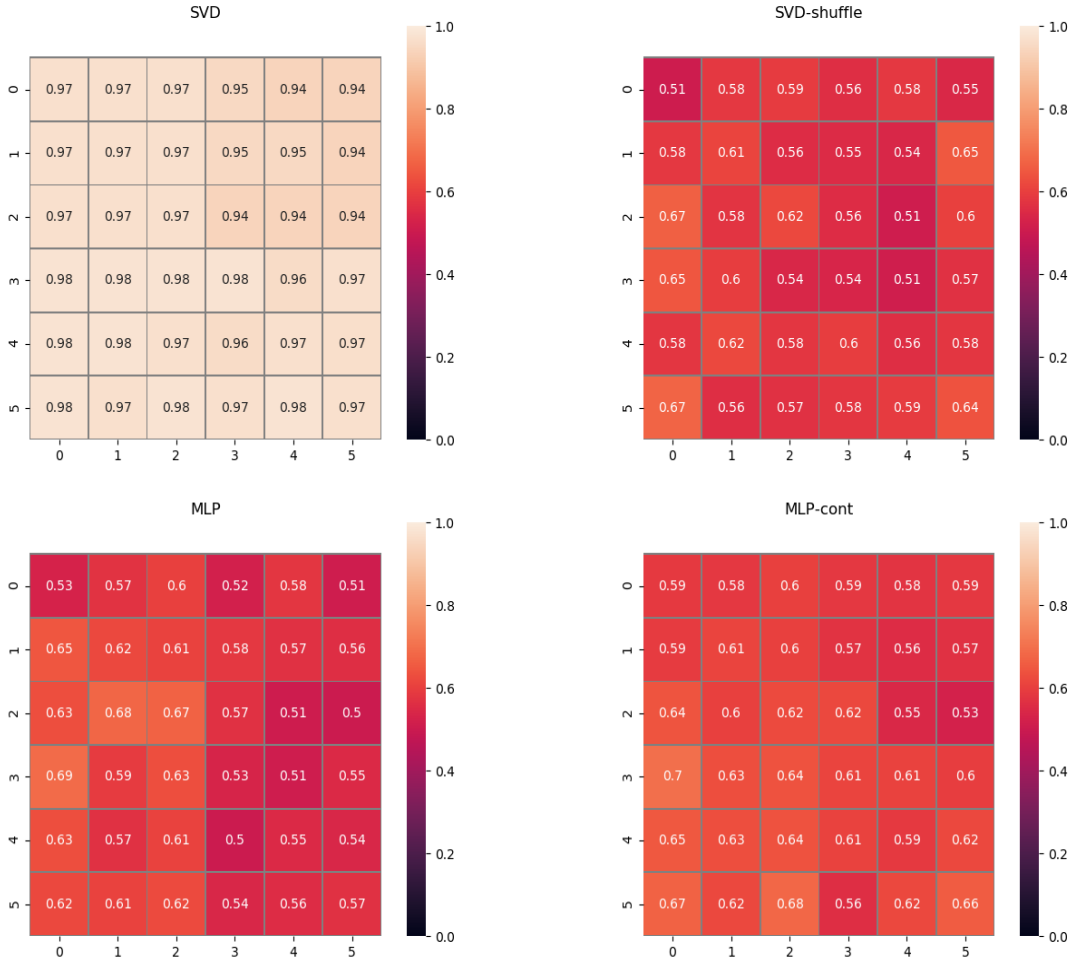
Control 2: Quadrants task

- The aims are to extend these findings to: 1) More complicated recombinations of the elementary tasks (richer transfer) 2) Non-exhaustive sampling in the elementary tasks
- Elementary tasks are categorization tasks between two quadrants. In total there are 4 choose 2 = 6 combinations of quadrants, and so there are 6 elementary tasks.
- Integration task is the same and still uses 8 random exemplars.

Quadrants, random order, test k=8

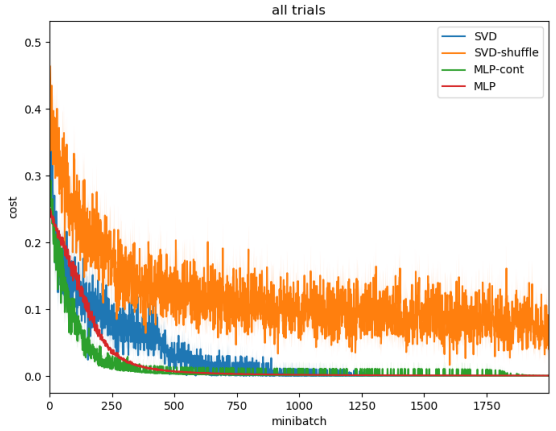


Mean accuracies, test k=8

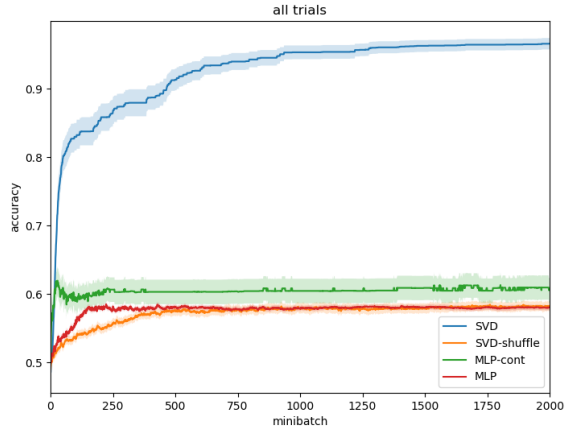


Temporal dynamics

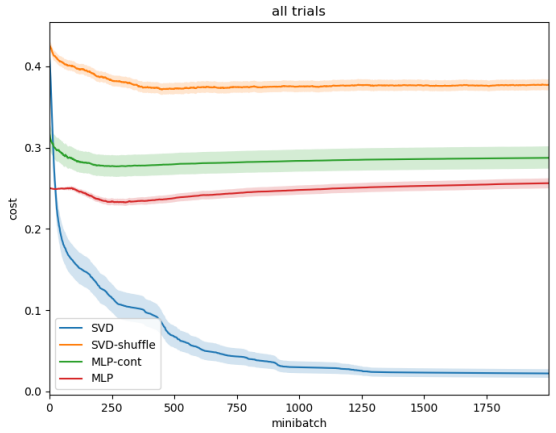
Integration training loss



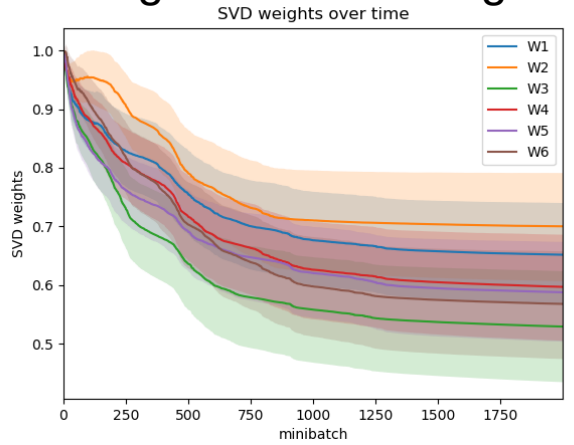
Integration test accuracy



Integration test loss



Integration SVD weights



Key questions

- How to model human-like compositional cognition?
- What are the benefits?
  - Few-shot learning of new categorization rules?
  - Extrapolation to unobserved regions of parameter space at test?

- Next steps?
  - Behavioral extrapolation experiment: model suggests that blocked training on orthogonal input features may facilitate few-shot learning and extrapolation in a subsequent integration task
  - Weather prediction task: curriculum learning of a compositional task (next part of presentation)
  - Modelling: ..?
  - Other suggestions most welcome!

Discussion

1) Compositionality does not only exist at the input feature level. Understanding of abstract properties and rules is likely built up in a similar manner across diverse experiences. An example of this would be transitivity in Steph's manipulations.

2) Handcrafting compositional features has been shown to aid learning of novel categories and generalization (Tokmakov et al., 2019), but extracting the underlying generative structure and mapping this back onto category membership in an automated, on-line fashion is an open problem.

3) Compression-recombination model relates to curriculum learning. Clustering inputs in sensible blocks facilitates learning, perhaps by using temporal proximity as a scaffold for extracting structure.