Universality and Individuality in Recurrent Neural Networks

Niru Maheswaranathan, Alex Williams, Matthew D Golub, Surya Ganguli, David Sussillo

Google Brain & Stanford University

NeurIPS 2019

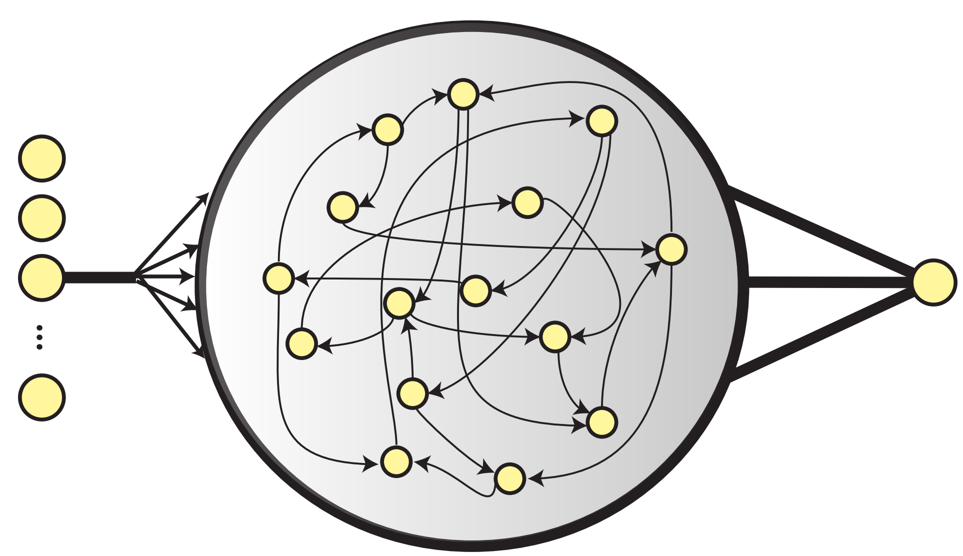
Poster #179

arxiv:1907.08549

Artificial neural networks in neuroscience

Advantages:

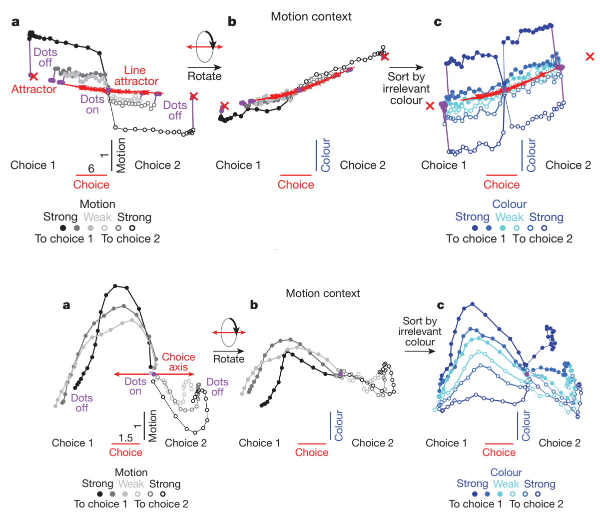
- Can train ANNs to accomplish tasks analogous to those studied in animals.
- Can inspect/probe/dissect artificial networks very easily.
- Can easily initiate a huge number of *in silico* studies



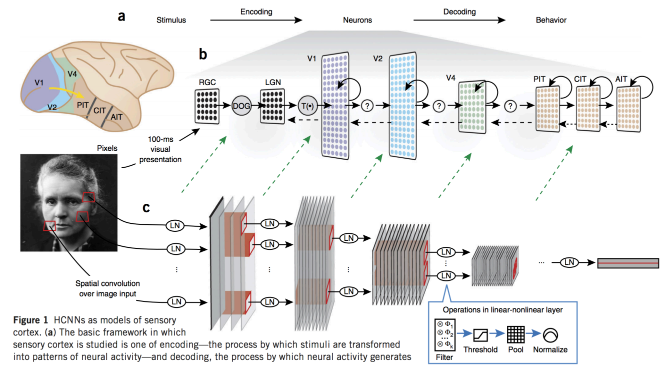
Poster #179

arxiv:1907.08549

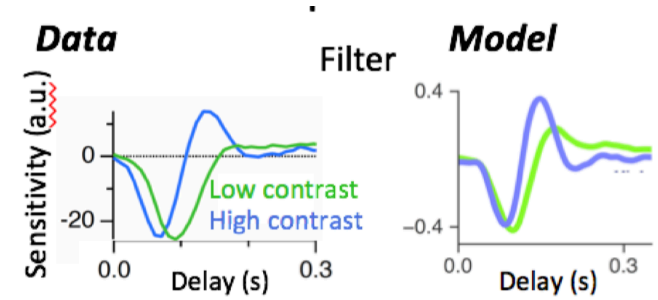
Artificial & biological neural networks



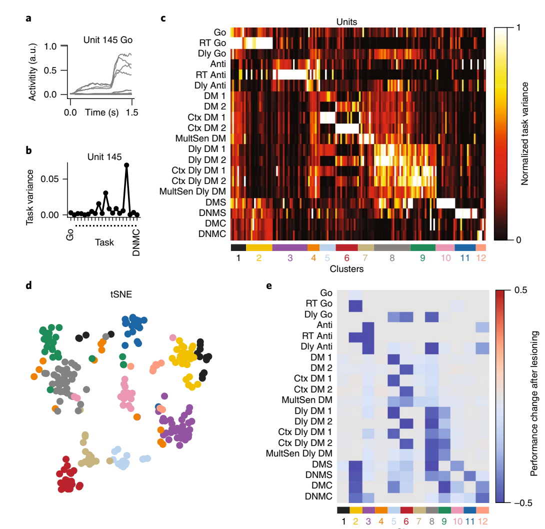
Mante & Sussillo et al. *Nature* 2013

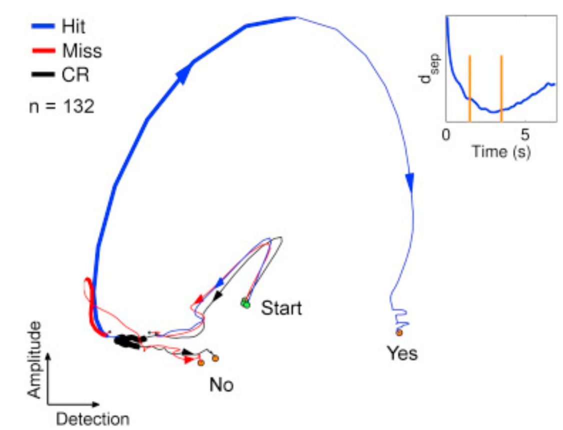


Yamins & DiCarlo, 2014

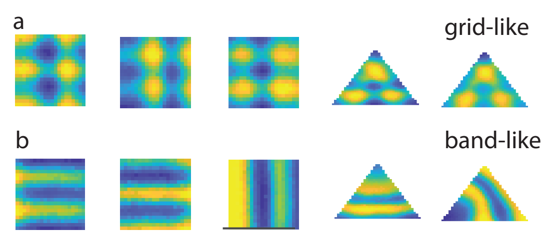


Maheswaranathan et al, 2018



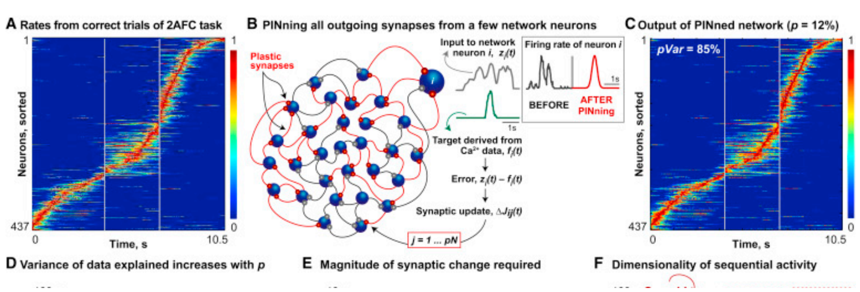


Carnevale et al, *NEURON*, 2015

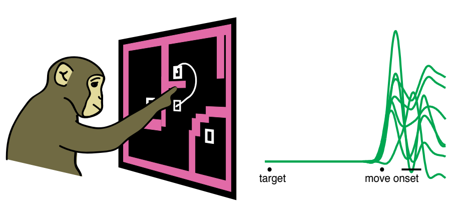


Cueva & Wei, *ICLR*, 2018

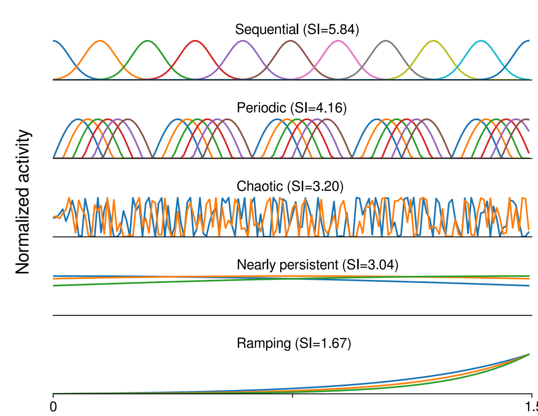
Yang, et al., *Nature Neuroscience* 2019



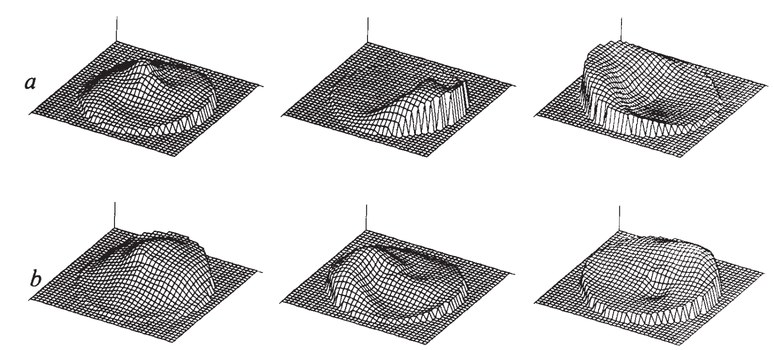
Rajan, Harvey, Tank, *Neuron*, 2016



Sussillo et al., *Nature Neuroscience*, 2015



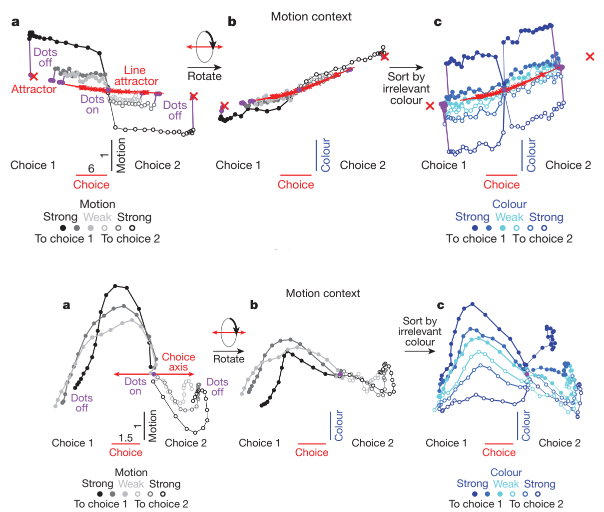
Orhan & Ma, *Nature Neuroscience*, 2019



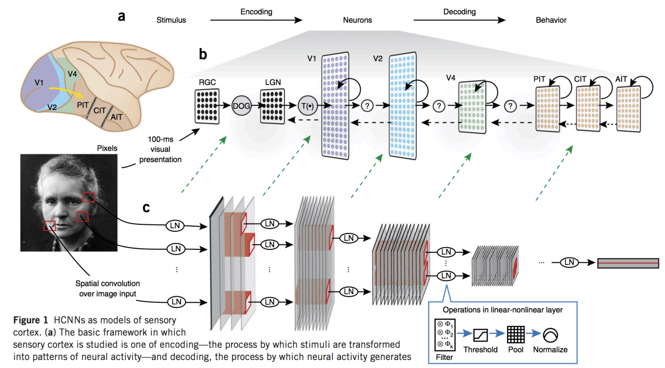
Zipser & Andersen, *Science*, 1988

Artificial & biological neural networks

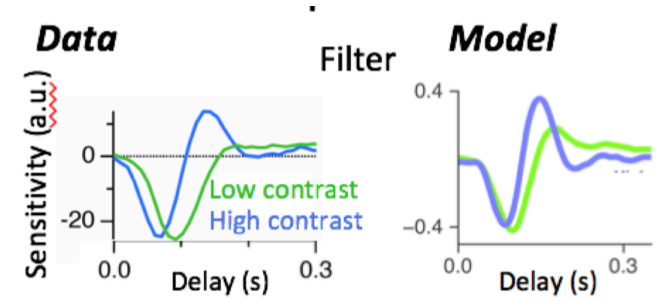
Networks have surprisingly similar representations…



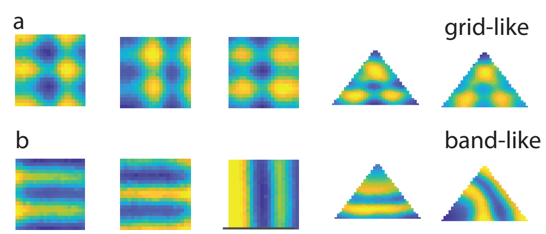
Mante & Sussillo et al. *Nature* 2013



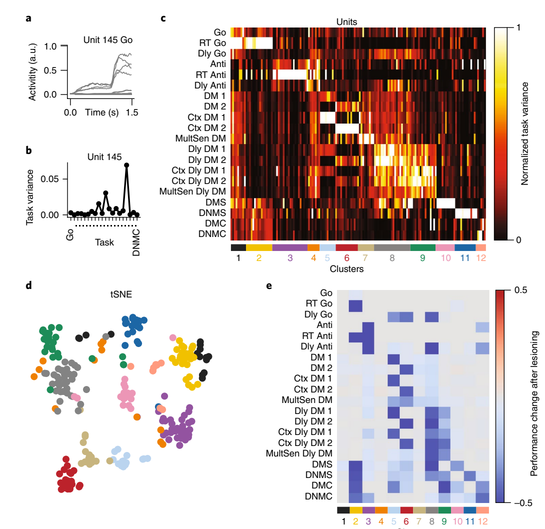
Yamins & DiCarlo, 2014



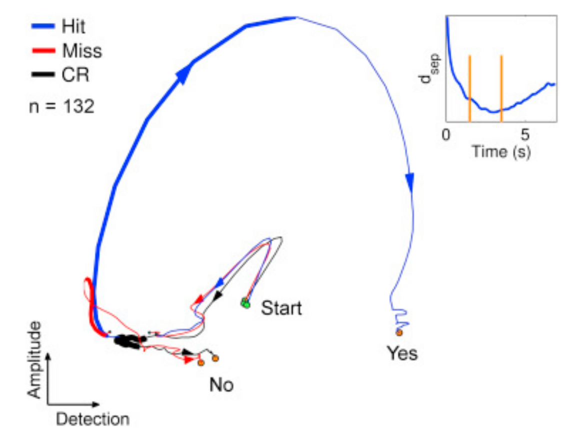
Maheswaranathan et al, 2018



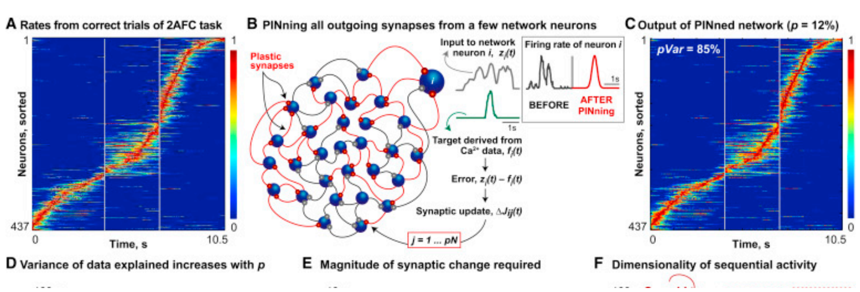
Cueva & Wei, *ICLR*, 2018



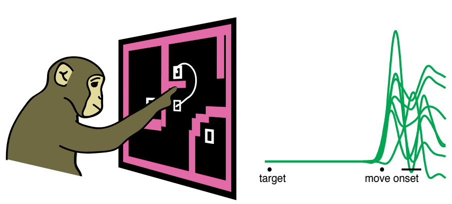
Yang, et al., *Nature Neuroscience* 2019



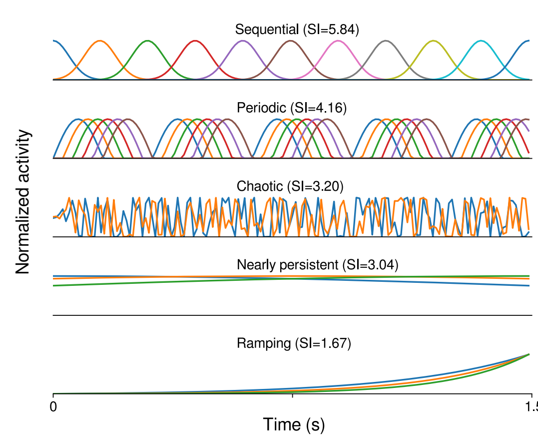
Carnevale et al, *NEURON*, 2015



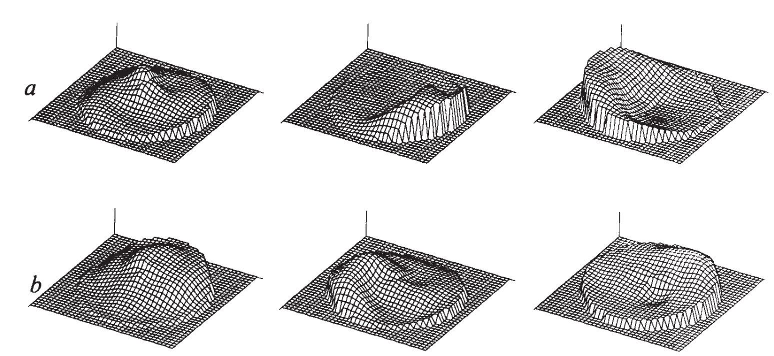
Rajan, Harvey, Tank, *Neuron*, 2016



Sussillo et al., *Nature Neuroscience*, 2015



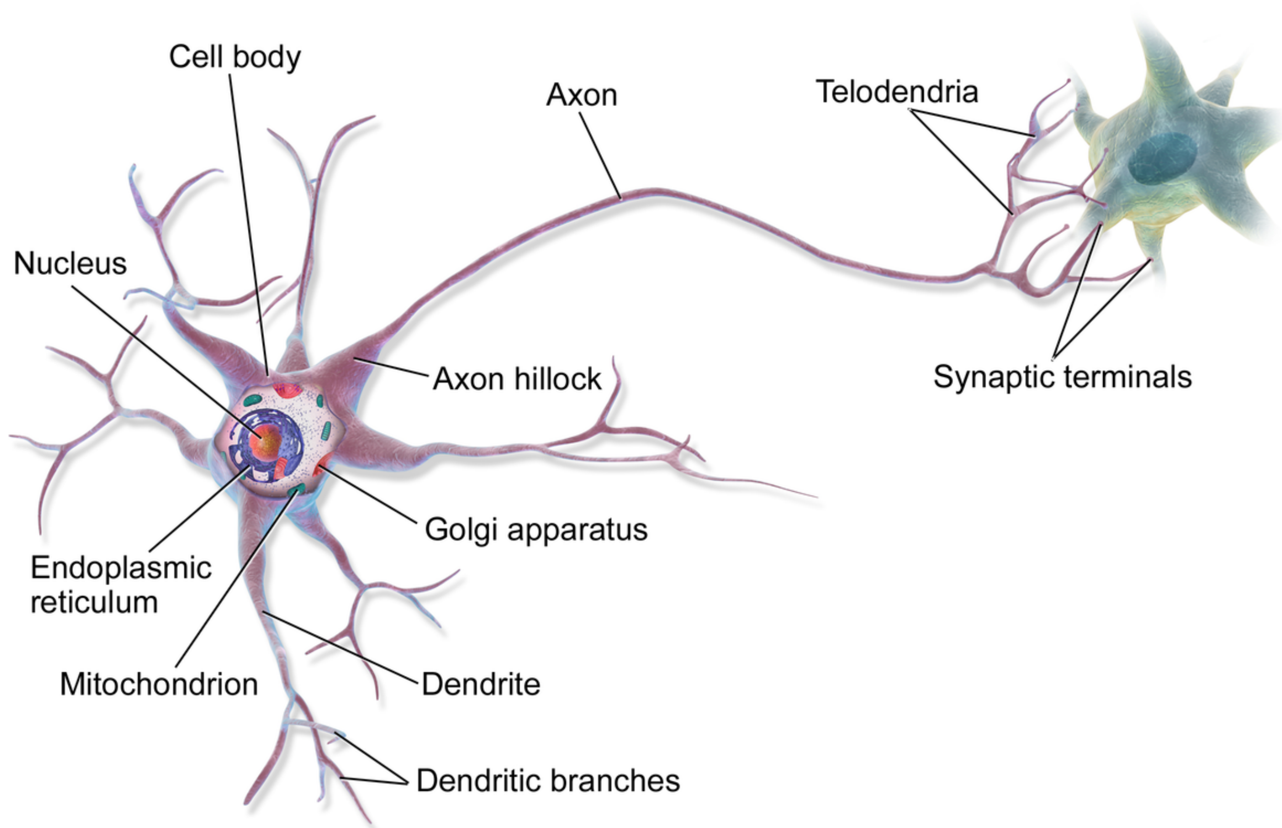
Orhan & Ma, *Nature Neuroscience*, 2019



Zipser & Andersen, *Science*, 1988

Artificial & biological neural networks

…but are composed of drastically different elements!



Biological neuron

$$y_i = \tanh(\sum_i W_{ij}\, x_j)$$

Artificial neuron

Poster #179

arxiv:1907.08549

Central question

When trained to perform the same **task**, why should we expect artificial and biological networks to be **similar**, given the drastic **differences in underlying mechanism**?

arxiv:1907.08549

This work: an empirical approach

Network mechanisms

RNN architectures (e.g. LSTMs, GRUs, …)
Nonlinearities (e.g. ReLU, tanh)
…

Similarity measures

Canonical correlation analysis (CCA)
Centered kernel alignment (CKA)

Tasks

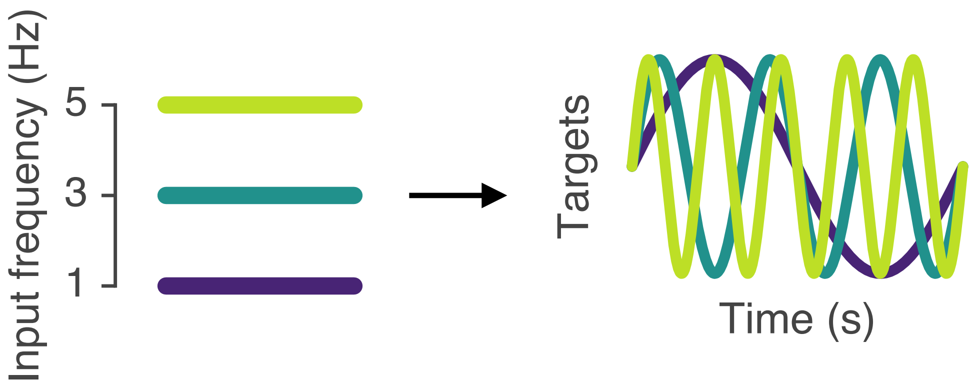
Decision making
Pattern generation
Working memory

arxiv:1907.08549

Evidence of both *universality* and *individuality*

Evidence of both *universality* and *individuality*

Pattern generation task

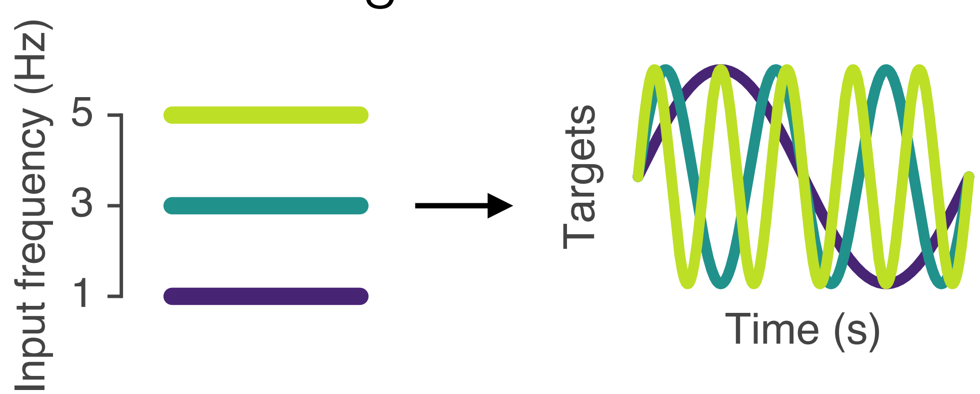


Poster #179

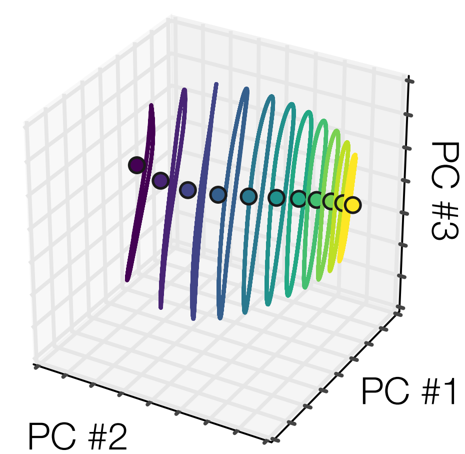
arxiv:1907.08549

Evidence of both *universality* and *individuality*

Pattern generation task



Analyzing trained networks

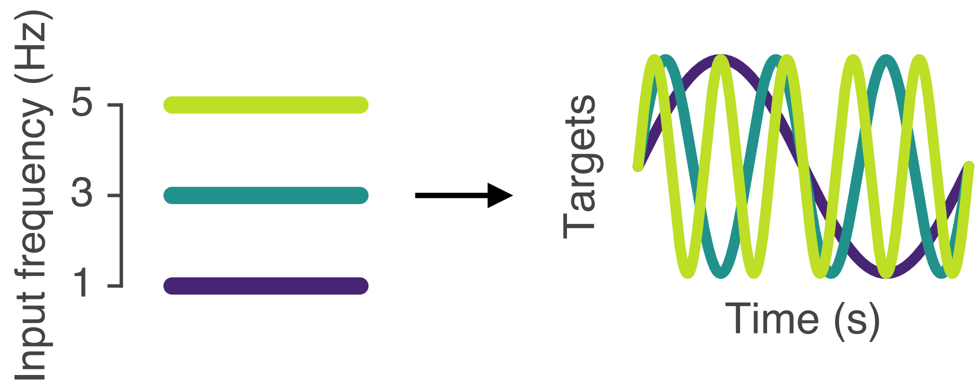


Poster #179

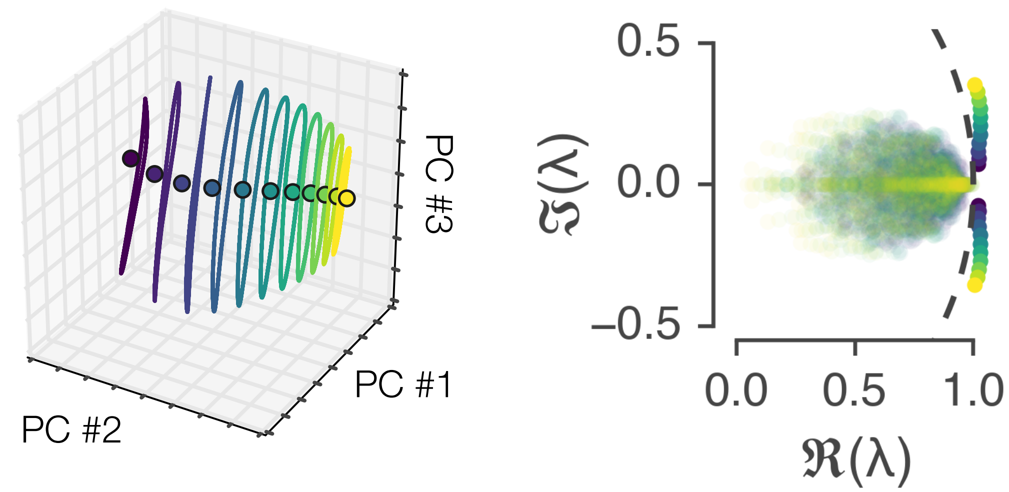
arxiv:1907.08549

Evidence of both *universality* and *individuality*

Pattern generation task



Analyzing trained networks

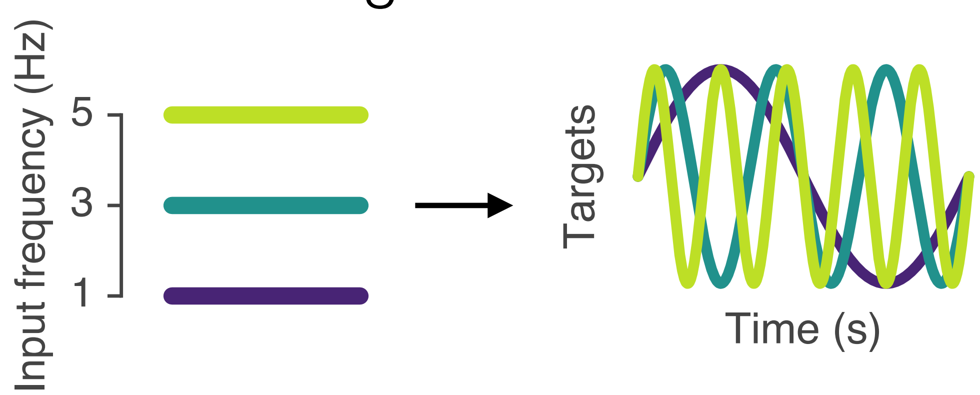


Poster #179

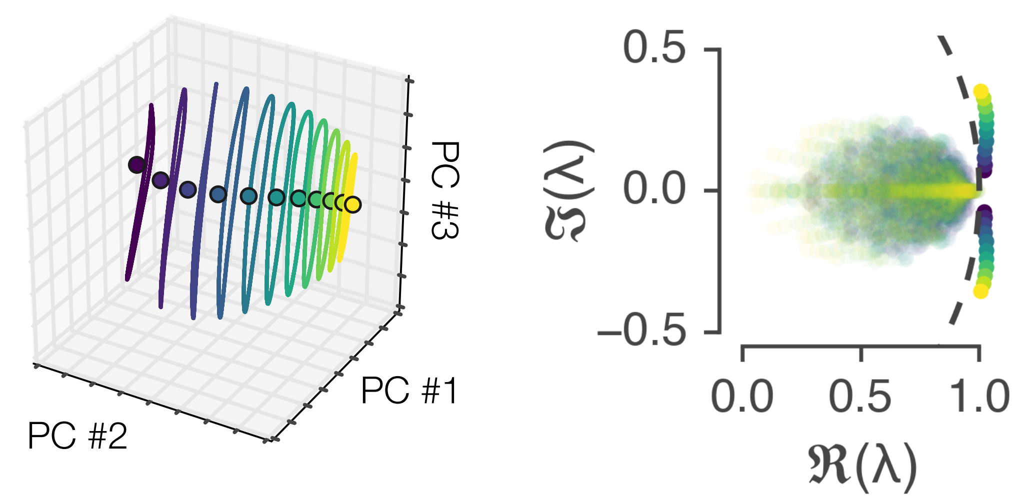
arxiv:1907.08549

Evidence of both *universality* and *individuality*

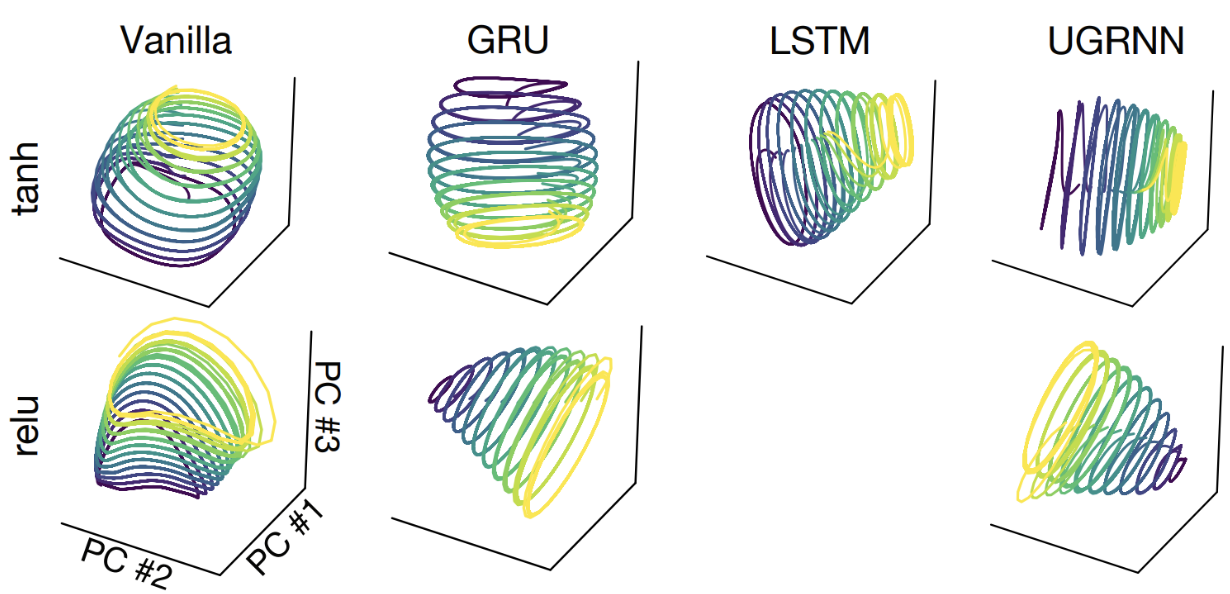
Pattern generation task



Analyzing trained networks



Network representations show individuality

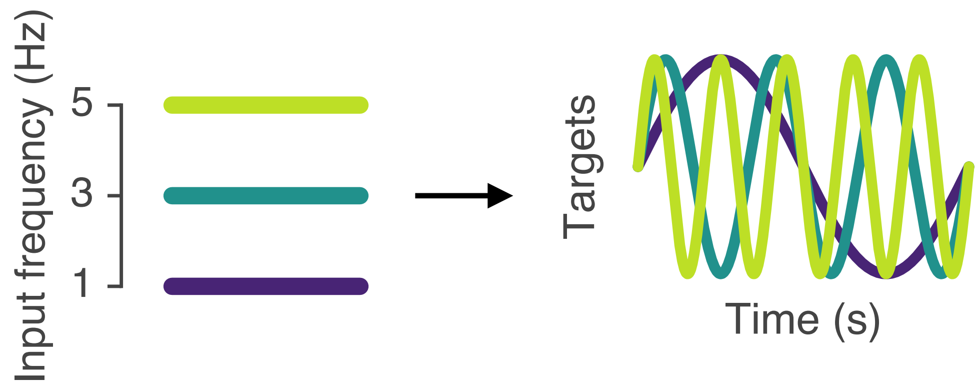


Poster #179

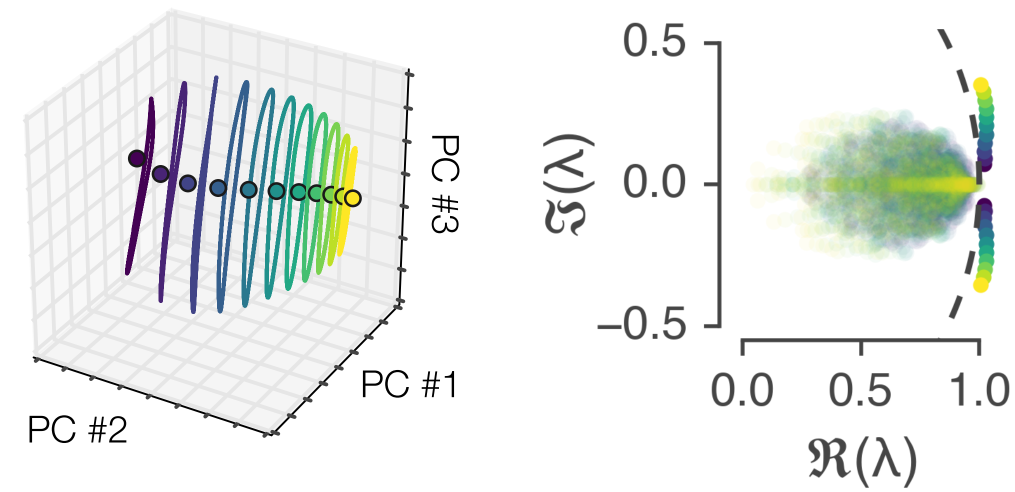
arxiv:1907.08549

Evidence of both *universality* and *individuality*

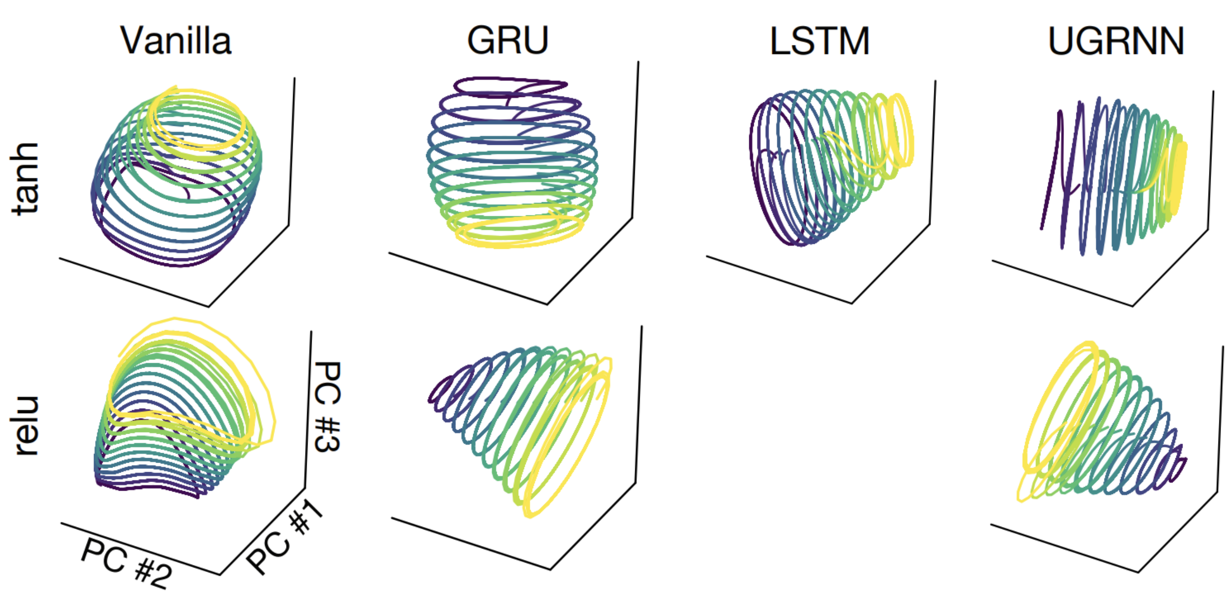
Pattern generation task



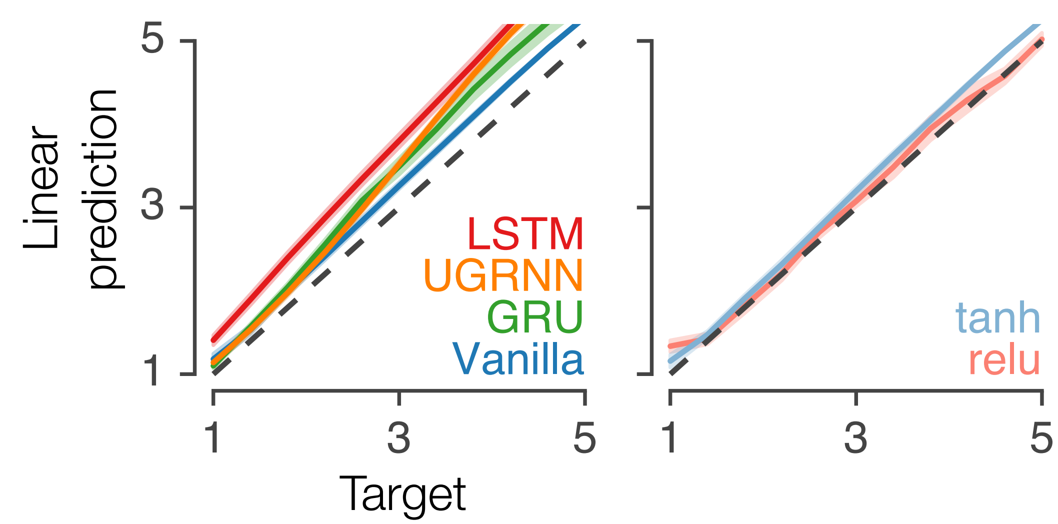
Analyzing trained networks



Network representations show individuality



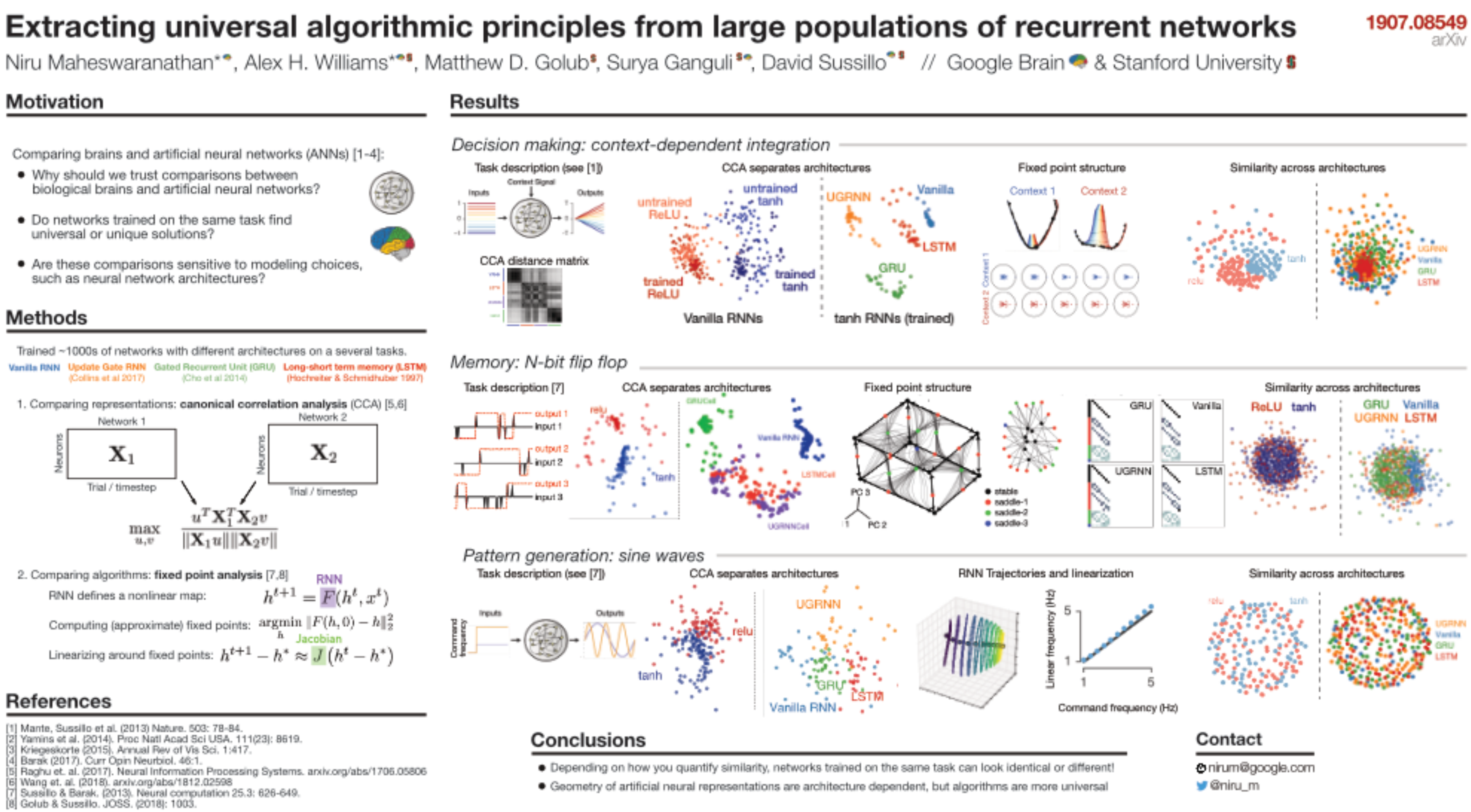
but aspects of the computation are universal



Poster #179

arxiv:1907.08549

Learn more at Poster #179



Poster #179

arxiv:1907.08549