Recurrent neural networks for neuroscience (Introduction)

Data Science and AI for Neuroscience Summer School, Caltech, July 15, 2022

Jonathan Kao

Assistant Professor
Dept of Electrical & Computer Engineering
University of California, Los Angeles

The deep learning revolution

https://image-net.org/challenges/LSVRC/2014/index.php

https://openai.com/blog/dall-e/

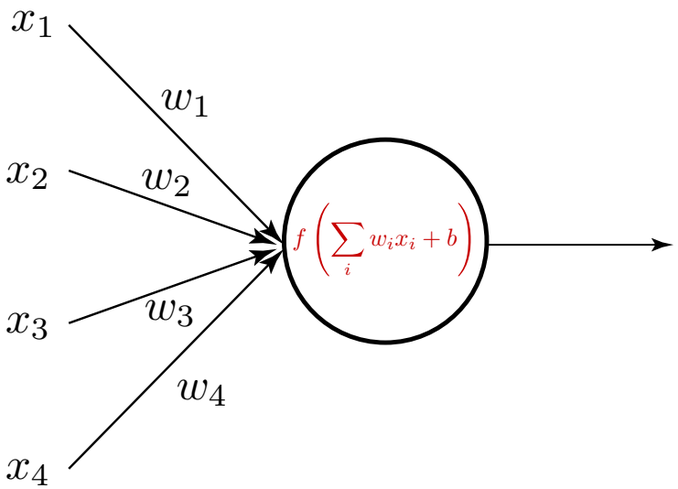
What if AI and humans do similar computations to perform tasks?

https://www.deepmind.com/research/highlighted-research/alphafold

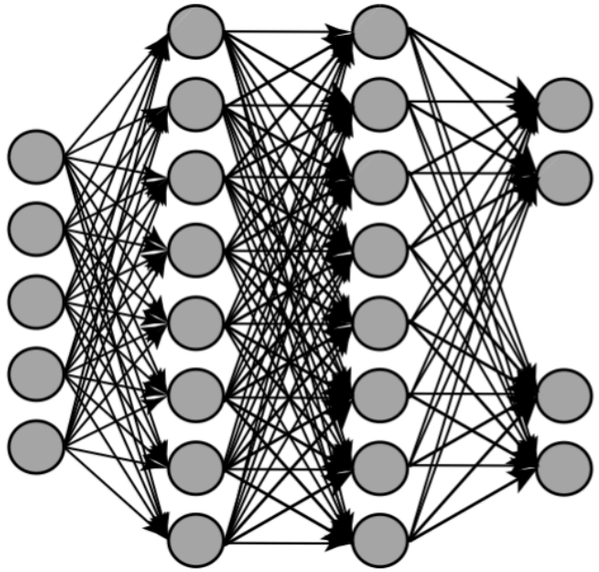
Game 3 of AlphaGo vs Lee Sedol, https://www.youtube.com/watch?v=_OV0Hlj8Fb8

Neural networks are precisely described by fully known equations

An artificial neuron

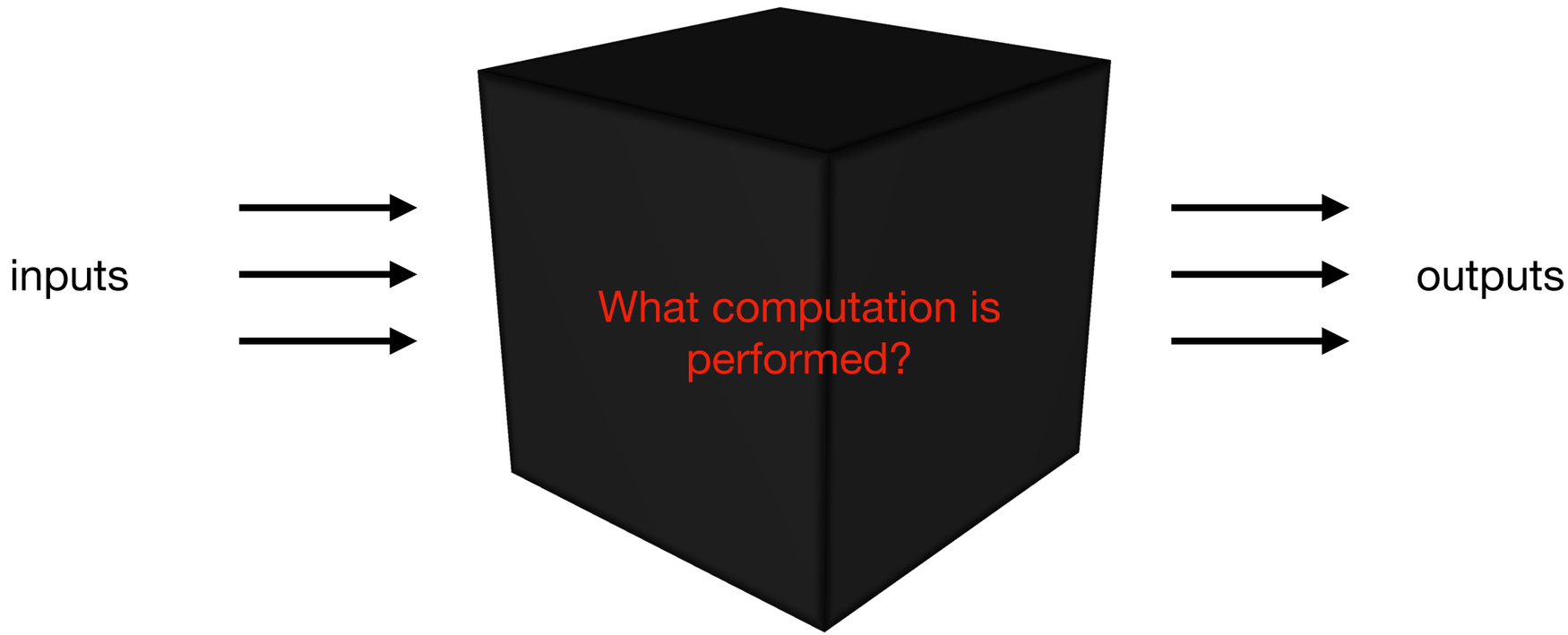


An artificial neural network



We even know the **learning rule** that sets the connection weights of these networks.

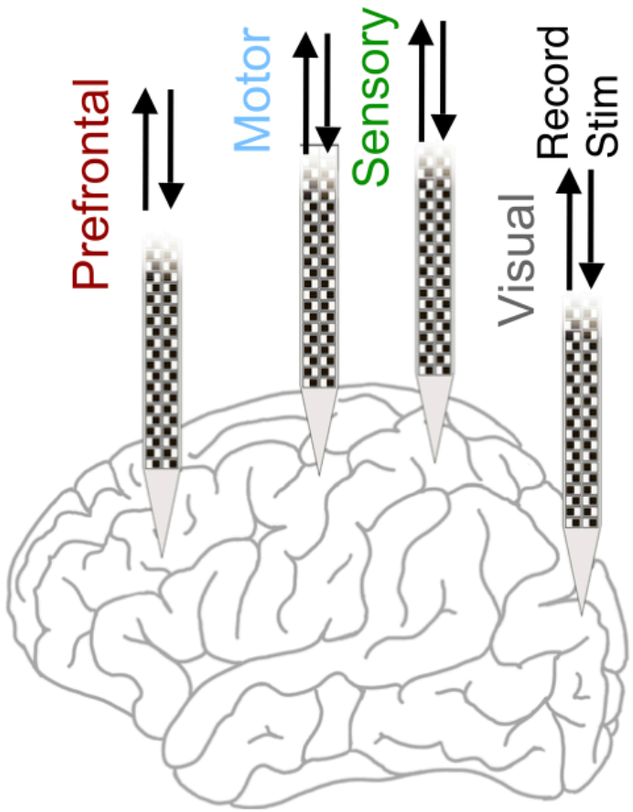
Neural networks are often treated as black boxes



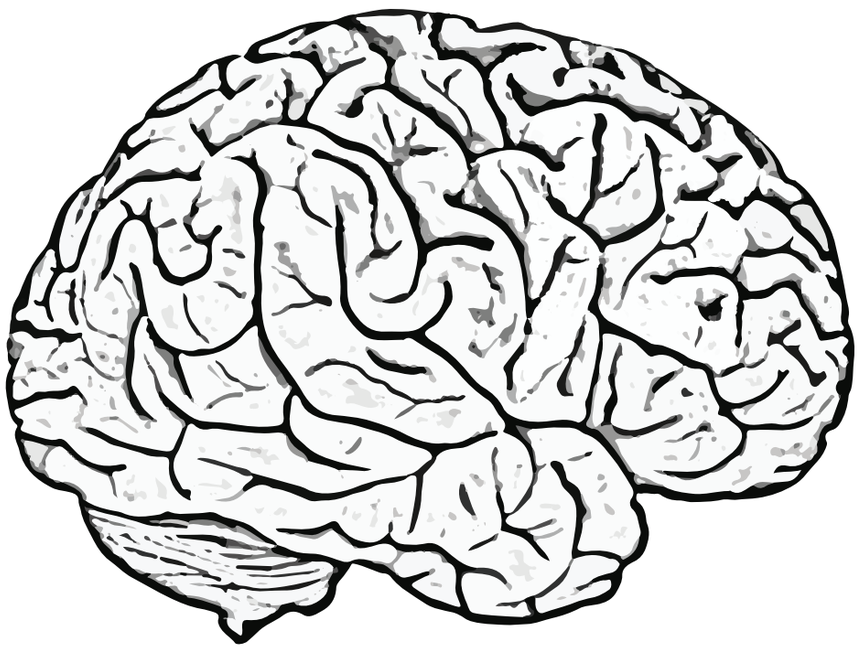
Feynman (and others): If you cannot explain something in simple terms, you don't understand it.

Back to the brain: more data, more insights?

Advancing neural technologies enable the simultaneous recording of hundreds to thousands of neurons across multiple brain areas.



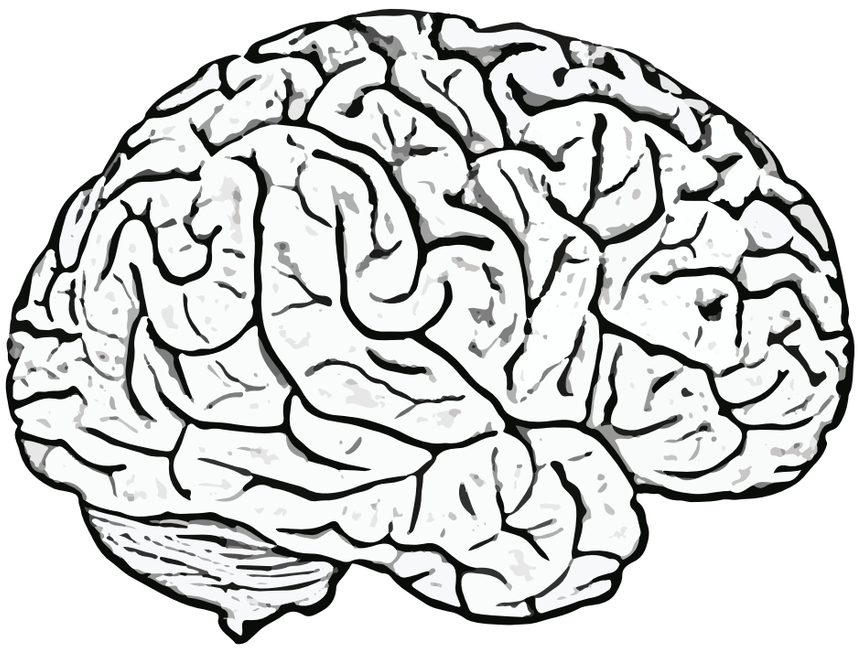
How do we understand the computation performed by neurons?



Some challenges / limitations

- **Sparse sampling:** we record at most hundreds to thousands of neurons.
- **Lack of connectivity:** we typically do not know how neurons are connected, let alone the synaptic weights.
- **Unobserved inputs:** we don’t observe the inputs to the neural circuit.
- **Limited learning rule knowledge:** the learning rule isn’t precisely known.

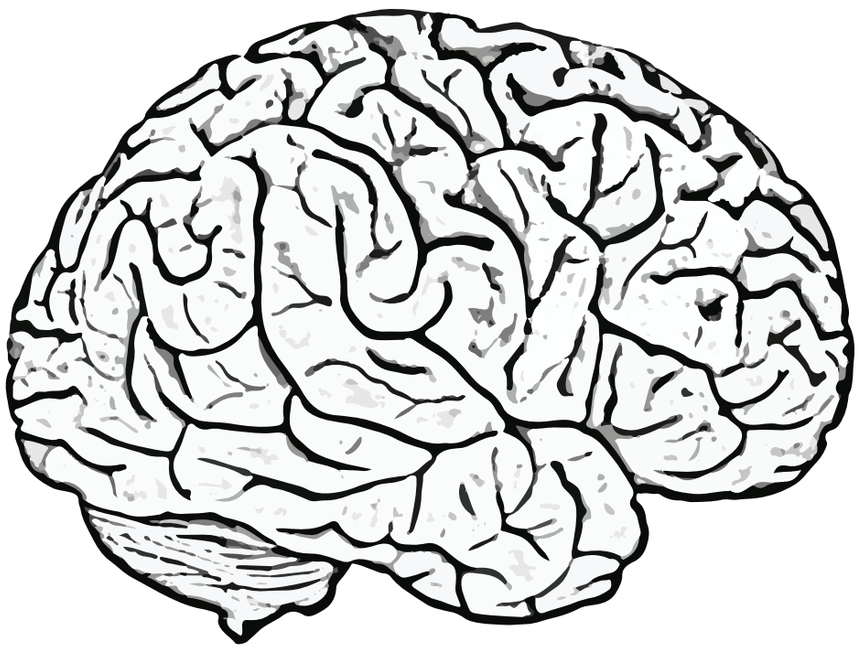
How do we understand the computation performed by neurons?



Aside: it is far from hopeless!

- We of course still learn a lot from this limited data. The entire neuroscience literature speaks to this.
- There is excellent progress in terms of understanding computation through other techniques, including dimensionality reduction and dynamical systems.
- **Critically, all these important approaches are complementary to what I'll present today.**

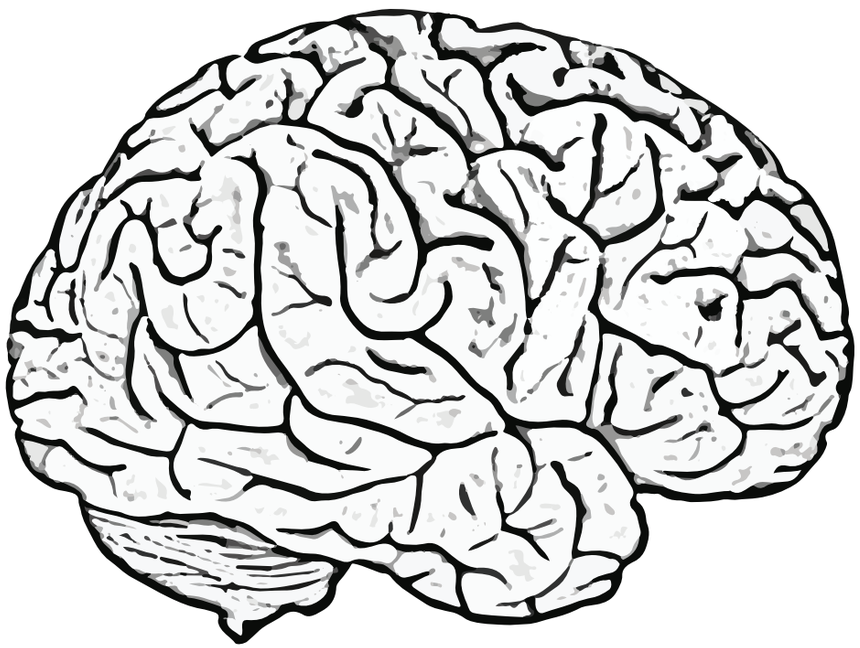
How do we understand the computation performed by neurons?



Some challenges / limitations

- **Sparse sampling:** we record at most hundreds to thousands of neurons.
- **Lack of connectivity:** we typically do not know how neurons are connected, let alone the synaptic weights.
- **Unobserved inputs:** we don’t observe the inputs to the neural circuit.
- **Limited learning rule knowledge:** the learning rule isn’t precisely known.

How do we understand the computation performed by neurons?



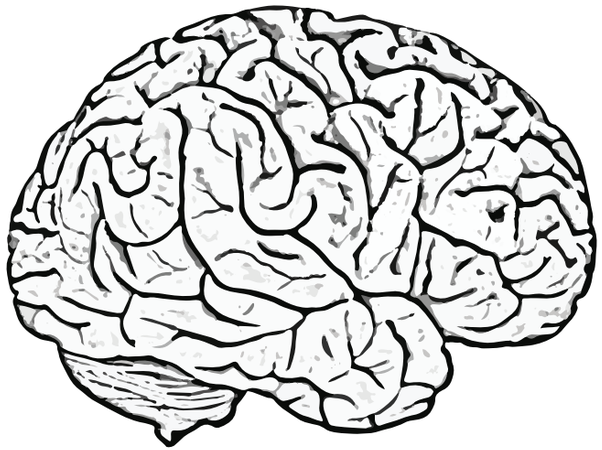
Some challenges / limitations

- **Sparse sampling:** we record at most hundreds to thousands of neurons.
- **Lacl** neurons are c
- **Unobserved inputs:** we don't observe the inputs to the neural circuit.
- **Limited learning rule knowledge:** the learning rule isn't precisely known.

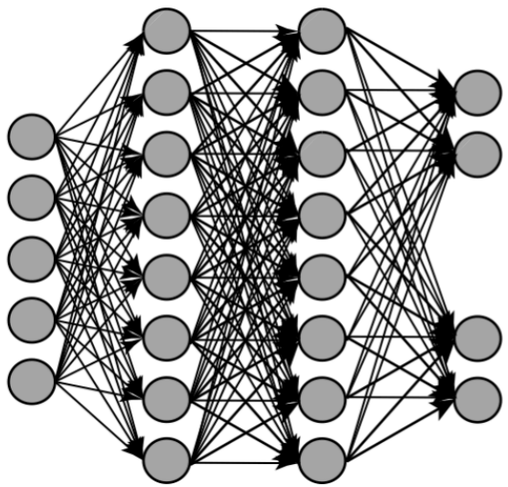
Can we make progress by studying a "simpler" system?

How do we understand the computation performed by neurons?

Cerebral cortex



Artificial neural network

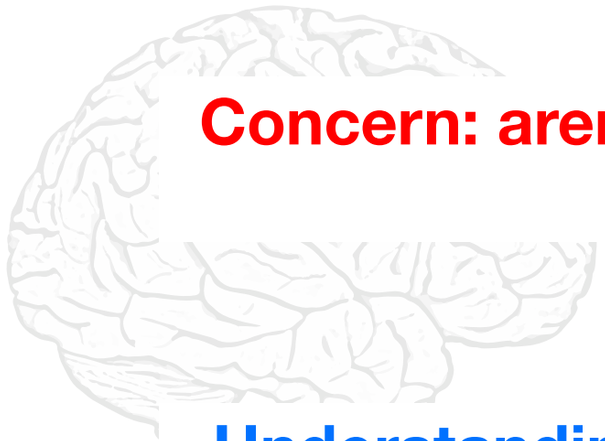


Some challenges / limitations

- ~~Sparse sampling.~~ Every artificial neuron observed.
- ~~Lack of connectivity.~~ Every connection known.
- ~~Unobserved inputs.~~ All inputs observed.
- ~~Limited learning rule knowledge.~~ Learning rule known.

What if we model a cortical computation with an ANN, then try to understand how the ANN does it?

How do we understand the computation performed by neurons?



Concern: aren't neural networks, though fully observed, too complex to understand?

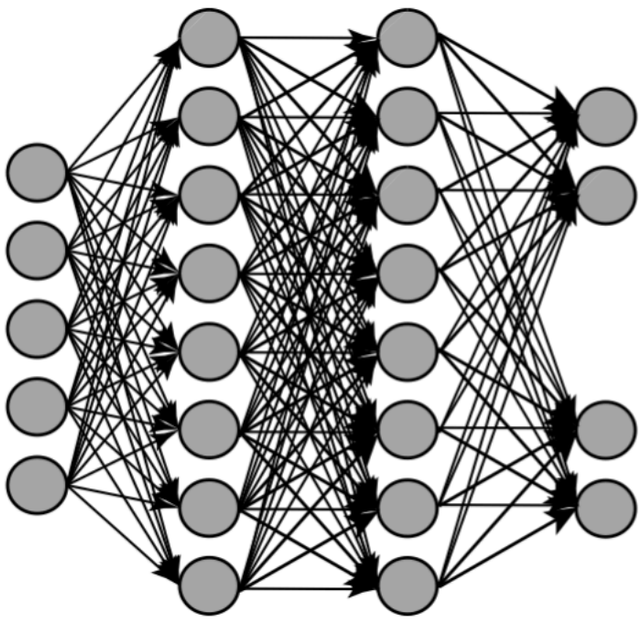
Understanding neural network computation is challenging. But if we can't understand a "simpler" artificial neural network's computation, won't it be much harder to understand the brain's computation?

And as we'll show, there has already been good progress.

Why recurrent neural networks?

RNNs aren’t necessary. In comparison to CNNs or other feedforward networks without recurrence,
RNNs have dynamics.

Feedforward NN (no recurrence, no dynamics)



Example:

Task: output next character in a word.

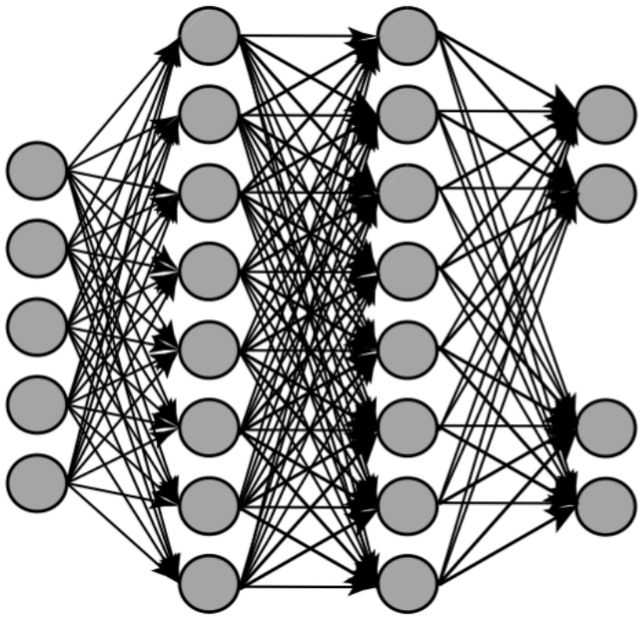
Input: the current letter in a string.
Output: the next letter in a string.

Trial 1: “Th” —> vowel
Trial 2: “Though” —> t

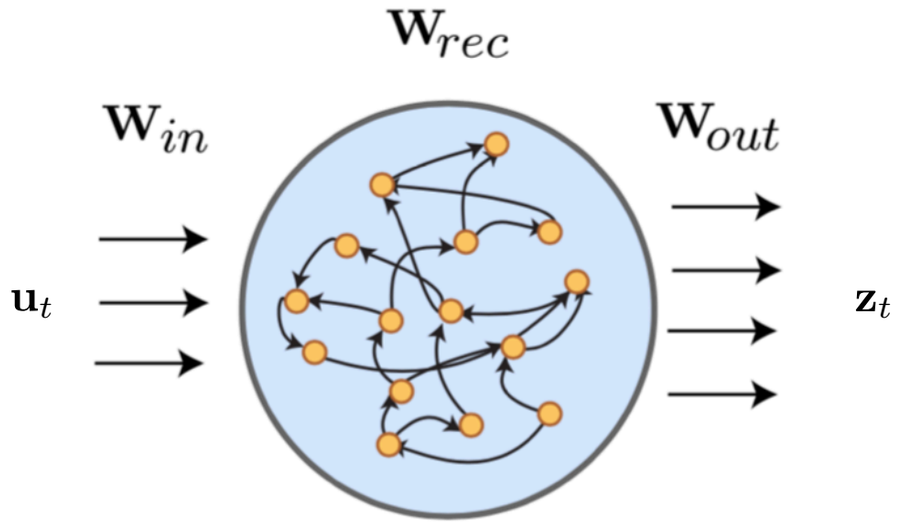
Why recurrent neural networks?

RNNs aren’t necessary. In comparison to CNNs or other feedforward networks without recurrence,
RNNs have dynamics.

Feedforward NN (no recurrence, no dynamics)

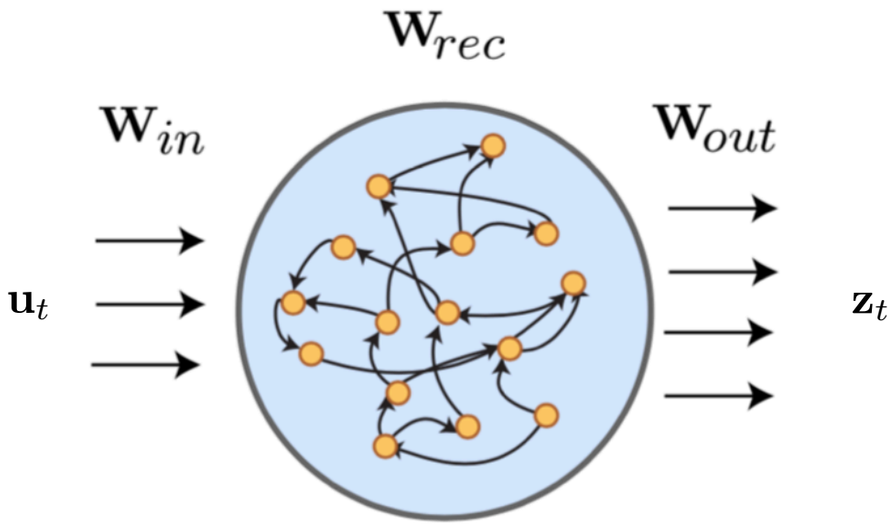


Feedback leads to dynamics



$$\tau\dot{\mathbf{x}}_t = -\mathbf{x}_t + \mathbf{W}_{\text{rec}} f(\mathbf{x}_t) + \mathbf{W}_{\text{in}}\mathbf{u}_t + \mathbf{b}$$

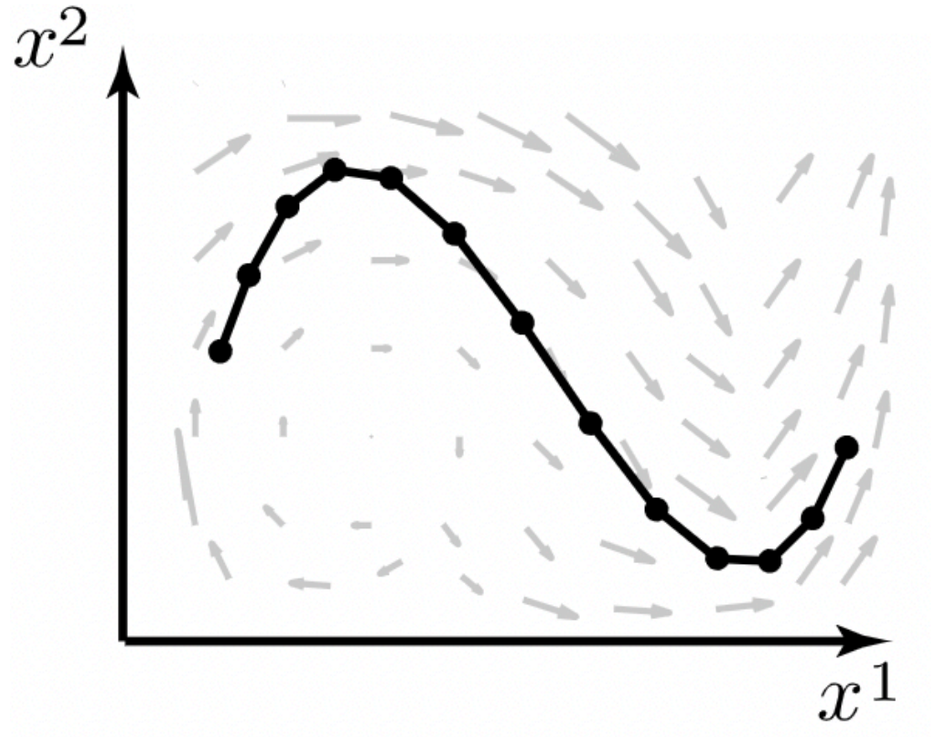
Why recurrent neural networks?



$$\tau \dot{\mathbf{x}}_t = -\mathbf{x}_t + \mathbf{W}_{\text{rec}} f(\mathbf{x}_t) + \mathbf{W}_{\text{in}} \mathbf{u}_t + \mathbf{b}$$

first order Euler appx

$$\mathbf{x}_{t+1} = (1-\alpha)\mathbf{x}_t + \alpha \left(\mathbf{W}_{\text{rec}} f(\mathbf{x}_t) + \mathbf{W}_{\text{in}} \mathbf{u}_t + \mathbf{b}_{\text{rec}} \right)$$



A recurrent neural network can approximate **any** dynamical system.

A basic way RNNs have been used for neuroscience

For one or multiple experimental tasks:

1. Define the task inputs ($\mathbf{u}_t$) and outputs ($\mathbf{z}_t$).

2. Train the RNN to learn task(s).

3. Assess if the *behavior* and *artificial neural activity* resemble empirical behavior and neural activity. If so, go to 4. If not, go back to 2 and modify the training.

4. Probe the RNN, leveraging its full observability, to develop hypotheses for the neural computation and predictions for future experiments.

A basic way RNNs have been used for neuroscience

For one or multiple experimental tasks:

1. Define the task inputs ($\mathbf{u}_t$) and outputs ($\mathbf{z}_t$).

2. Train the RNN to learn task(s).

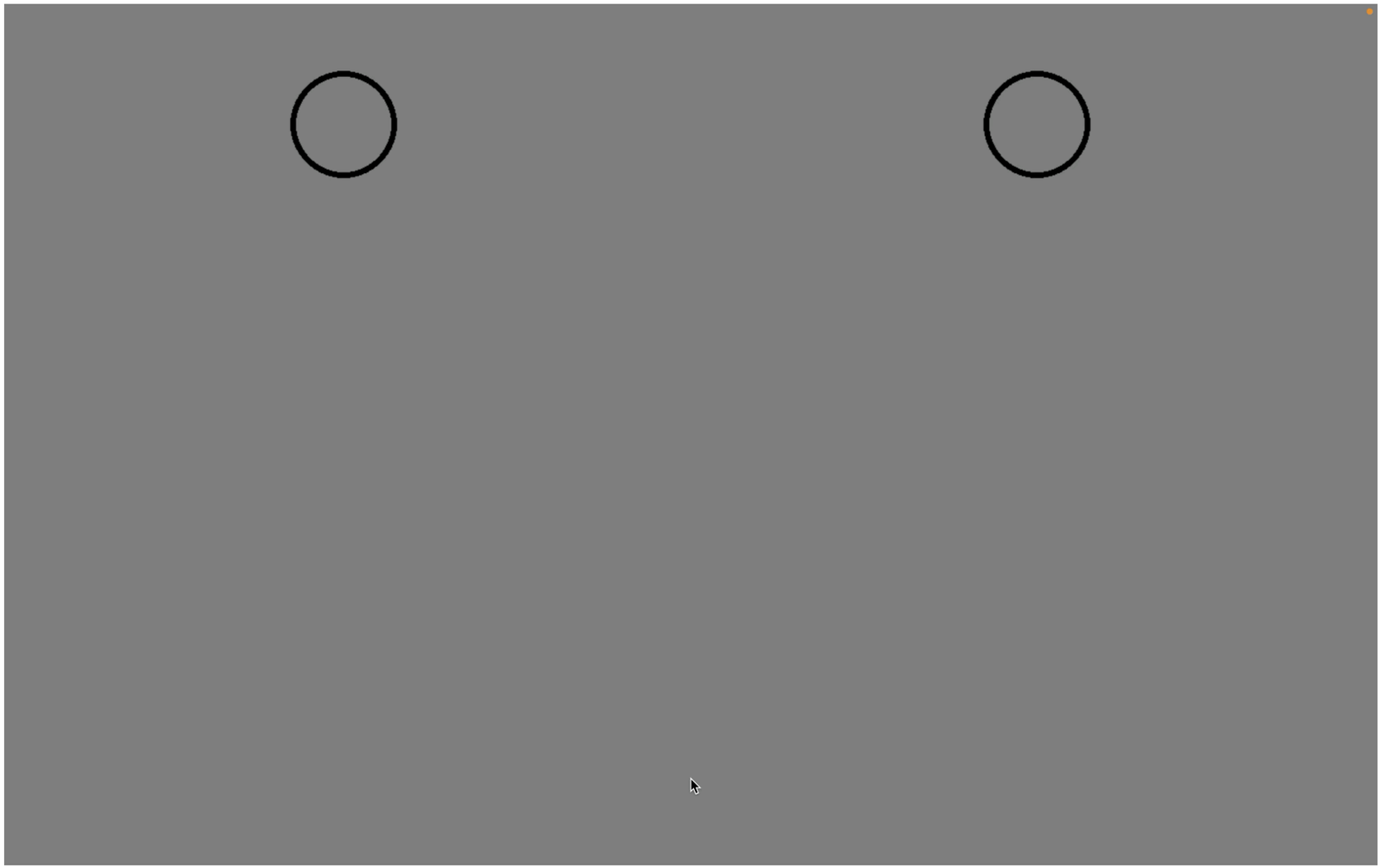
3. Assess if the *behavior* and *artificial neural activity* resemble empirical behavior and neural activity. If so, go to 4. If not, go back to 2 and modify the training.

4. Probe the RNN, leveraging its full observability, to develop hypotheses for the neural computation and predictions for future experiments.

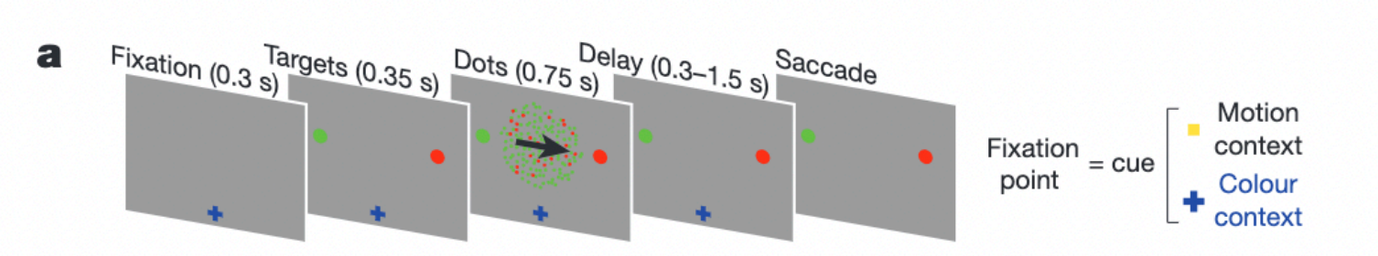
Examples include: A big literature here, with many fascinating implementations, and many who I can't reference for the sake of time. E.g., Sussillo, Abbott, Buonomano, Rajan, Ostojic, Ganguly, Yang, Hennequin, Jazayeri, Wang, Shenoy, Newsome, Barak, DeepMind, Yamins, DiCarlo, Michaels, Scherberger, Saxena, Cunningham, list goes on.

Example 1: Mante*, Sussillo*, et al., *Nature 2013*

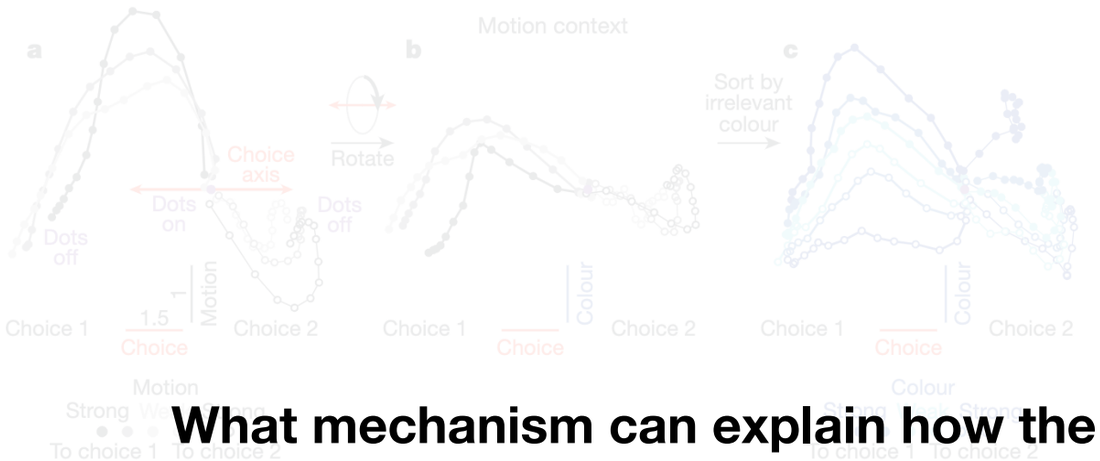
Brandon McMahan (MD/PhD student) and I prepared a colab notebook where you will implement an RNN, train it, and reproduce the results of this paper.



Example 1: Mante*, Sussillo*, et al., *Nature 2013*

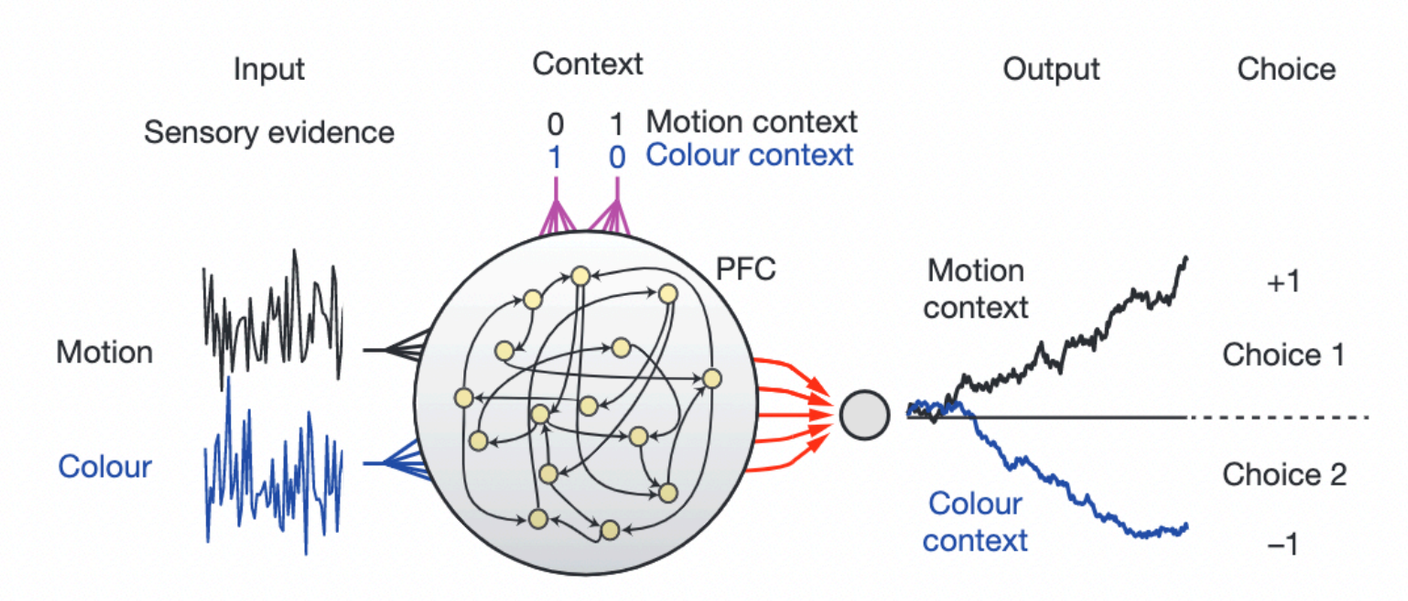


Example 1: Mante*, Sussillo*, et al., *Nature 2013*



What mechanism can explain how the brain does selective integration?

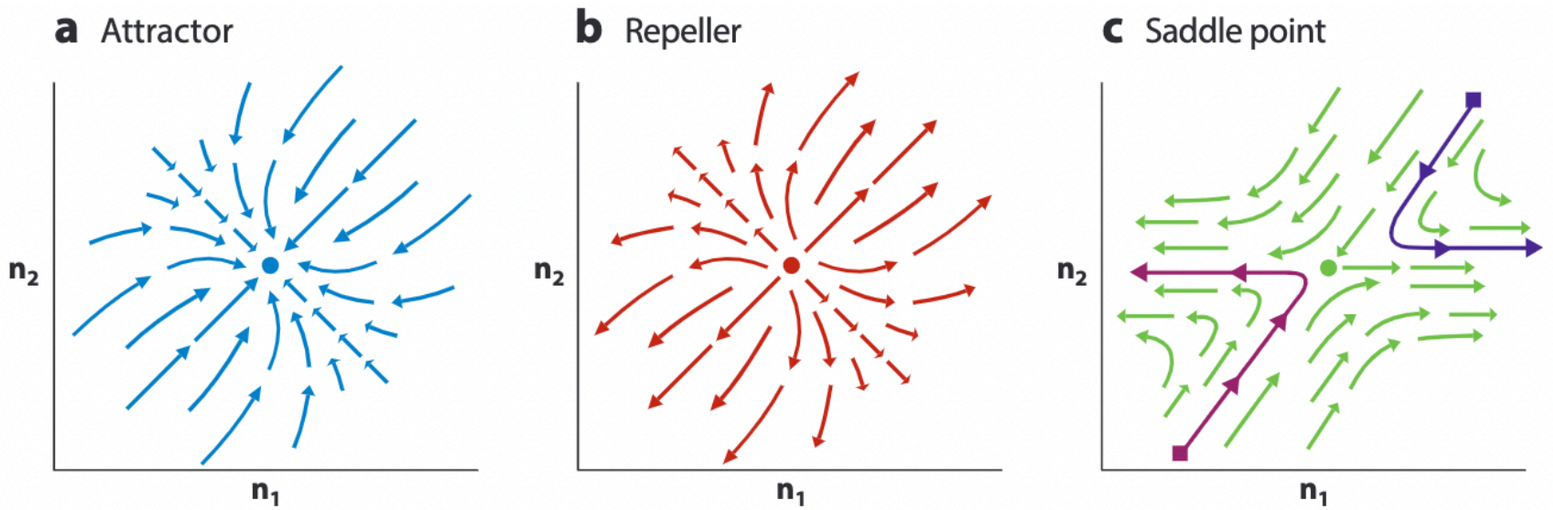
Example 1: Mante*, Sussillo*, et al., *Nature 2013*



Brief aside: Probing RNN dynamics

$$\tau \dot{\mathbf{x}}_t = -\mathbf{x}_t + \mathbf{W}_{\text{rec}} f(\mathbf{x}_t) + \mathbf{W}_{\text{in}} \mathbf{u}_t + \mathbf{b}$$

Solving this equation for $\dot{\mathbf{x}}_t = 0$ identifies the fixed points of the RNN dynamical system.

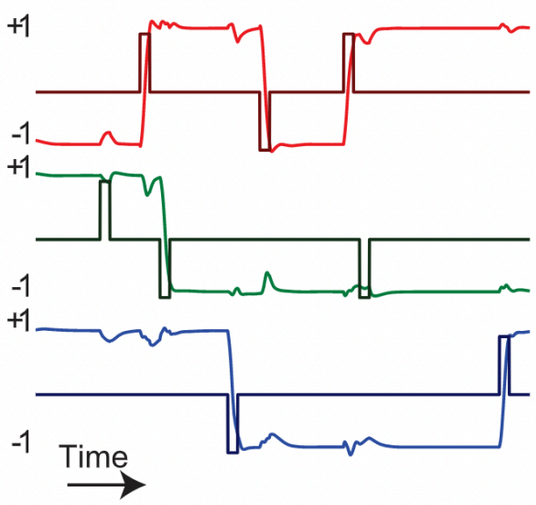


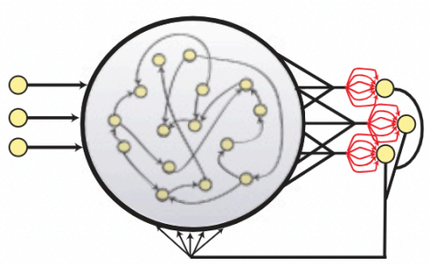
Vyas et al., ARN 2020

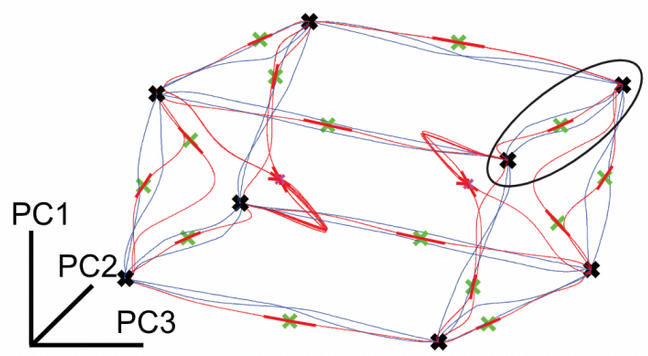
Brief aside: Probing RNN dynamics

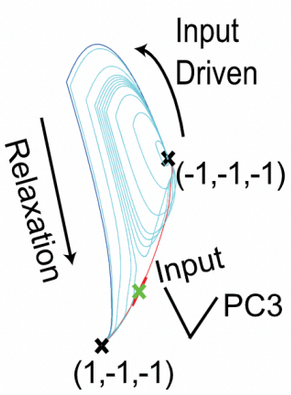
$$\tau \dot{\mathbf{x}}_t = -\mathbf{x}_t + \mathbf{W}_{\text{rec}} f(\mathbf{x}_t) + \mathbf{W}_{\text{in}} \mathbf{u}_t + \mathbf{b}$$

Solving this equation for $\dot{\mathbf{x}}_t = 0$ identifies the fixed points of the RNN dynamical system.







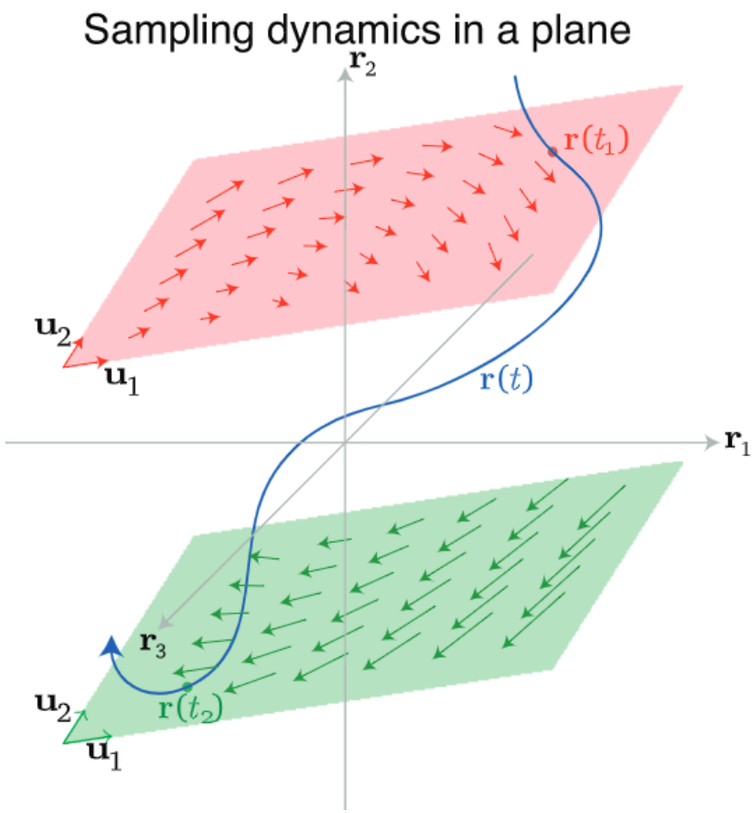


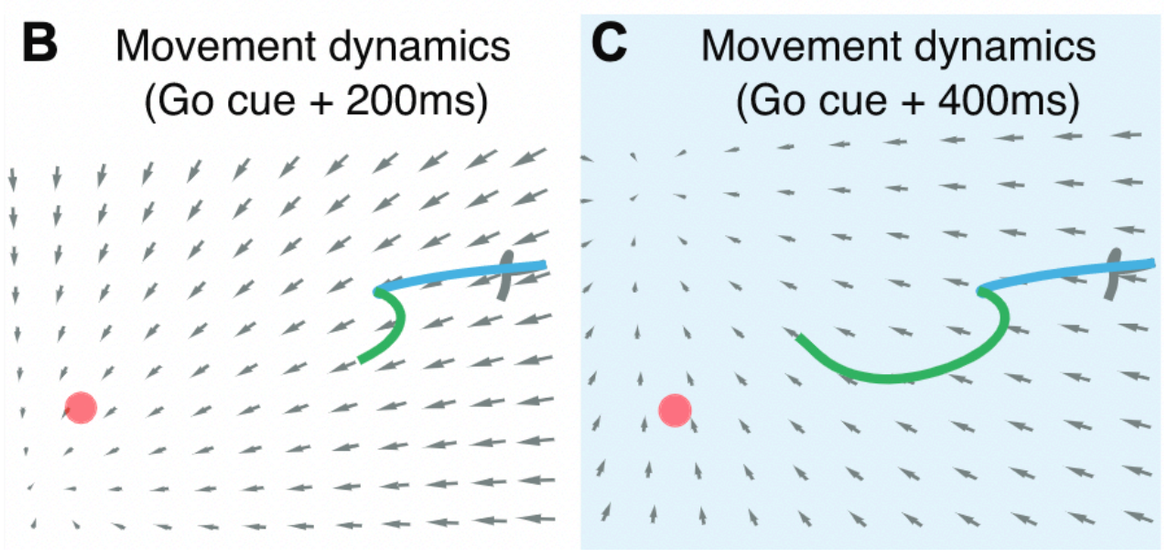
Sussillo & Barak, 2013

Brief aside: Probing RNN dynamics

$$\tau \dot{\mathbf{x}}_t = -\mathbf{x}_t + \mathbf{W}_{\text{rec}} f(\mathbf{x}_t) + \mathbf{W}_{\text{in}} \mathbf{u}_t + \mathbf{b}$$

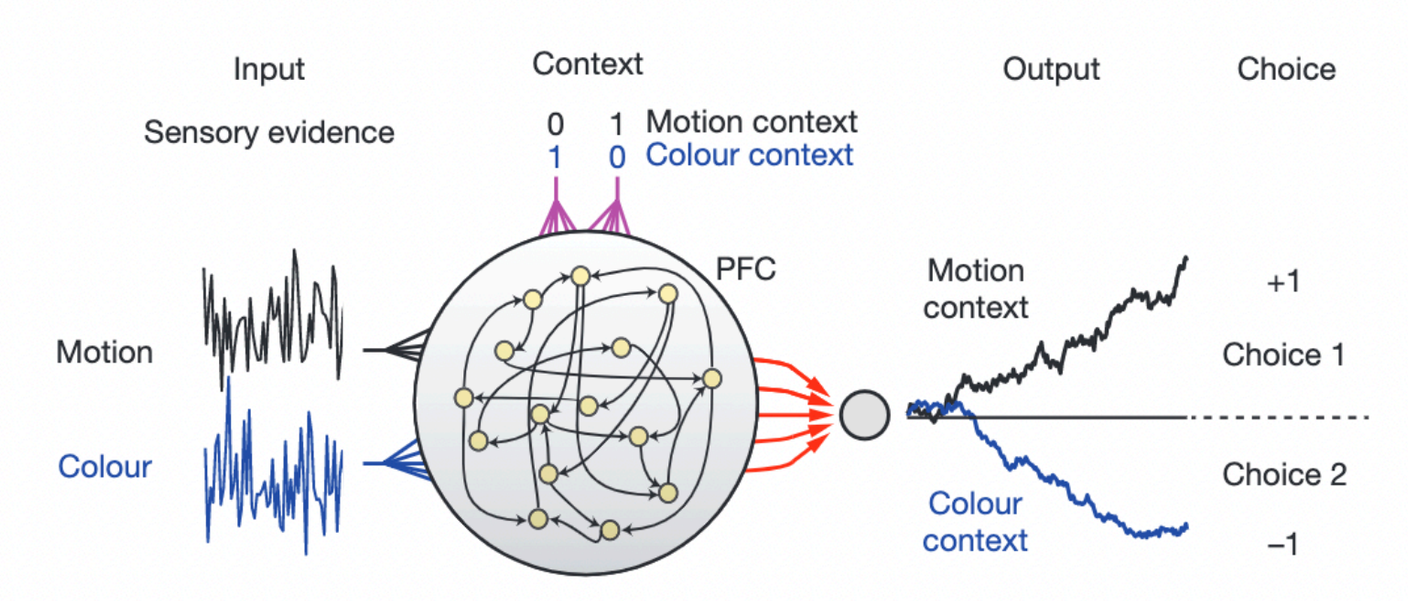
Projecting $\dot{\mathbf{x}}_t$ into a plane allows visualizing dynamics (though one should be careful doing this)!



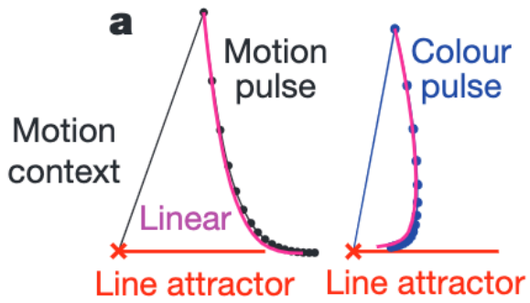


Kao, 2019

Example 1: Mante*, Sussillo*, et al., *Nature 2013*



Example 1: Mante*, Sussillo*, et al., *Nature 2013*



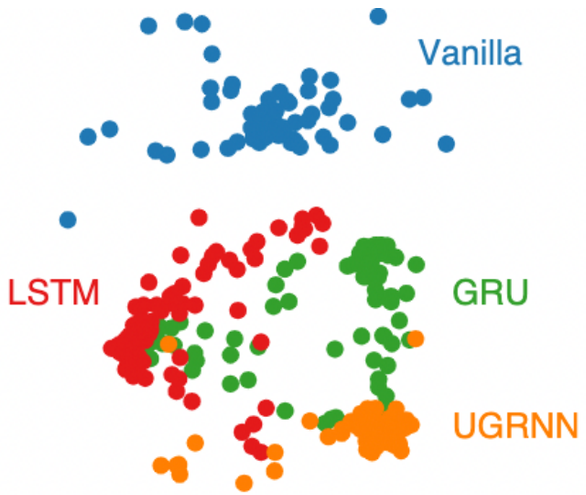
This is a new mechanism for selective integration.

A relevant question

How robust are the representations and dynamical mechanisms to architecture, nonlinearity, and learning rule?

Similarity of representations

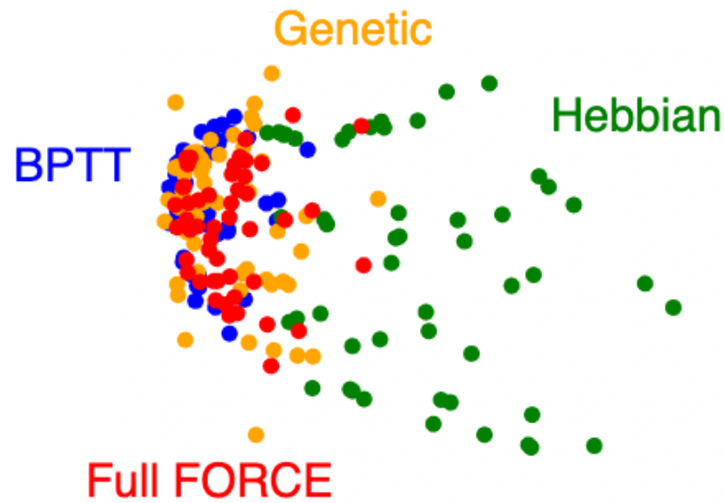
Architectures



Maheswaranathan, Williams et al., NeurIPS 2019

Nonlinearity

Learning rule



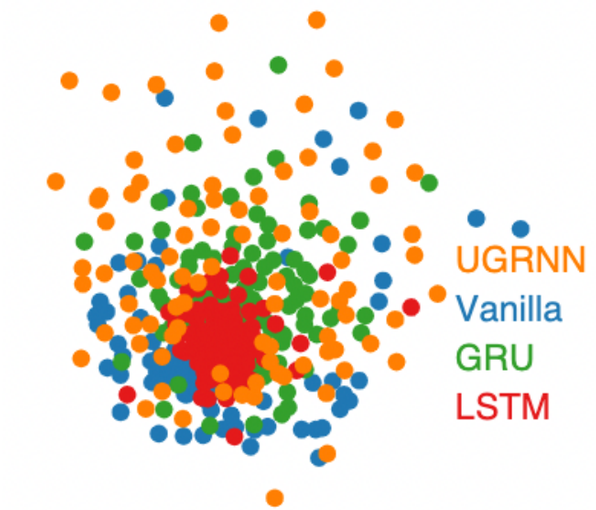
McMahan, Kleinman, Kao, NeurIPS 2021

A relevant question

How robust are the representations and dynamical mechanisms to architecture, nonlinearity, and learning rule?

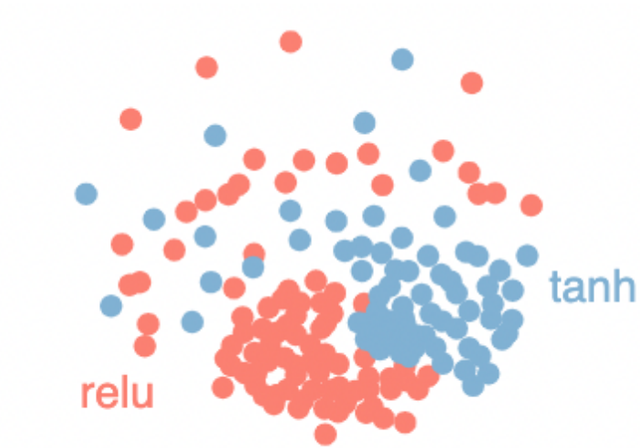
Similarity of dynamics

Architectures

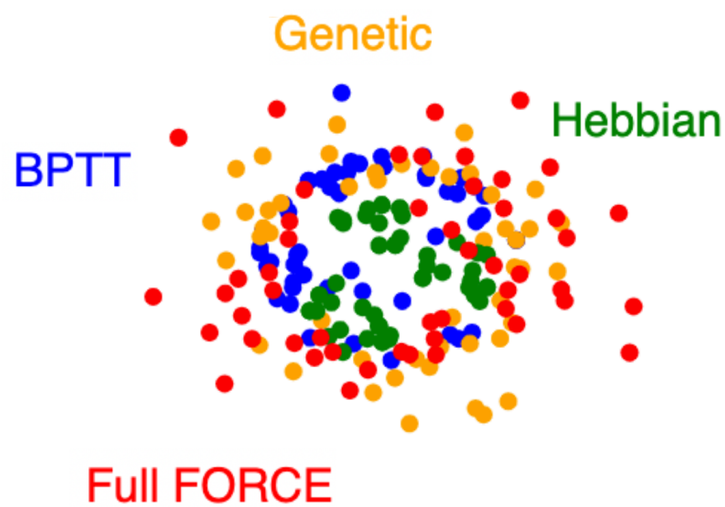


Maheswaranathan, Williams et al., NeurIPS 2019

Nonlinearity



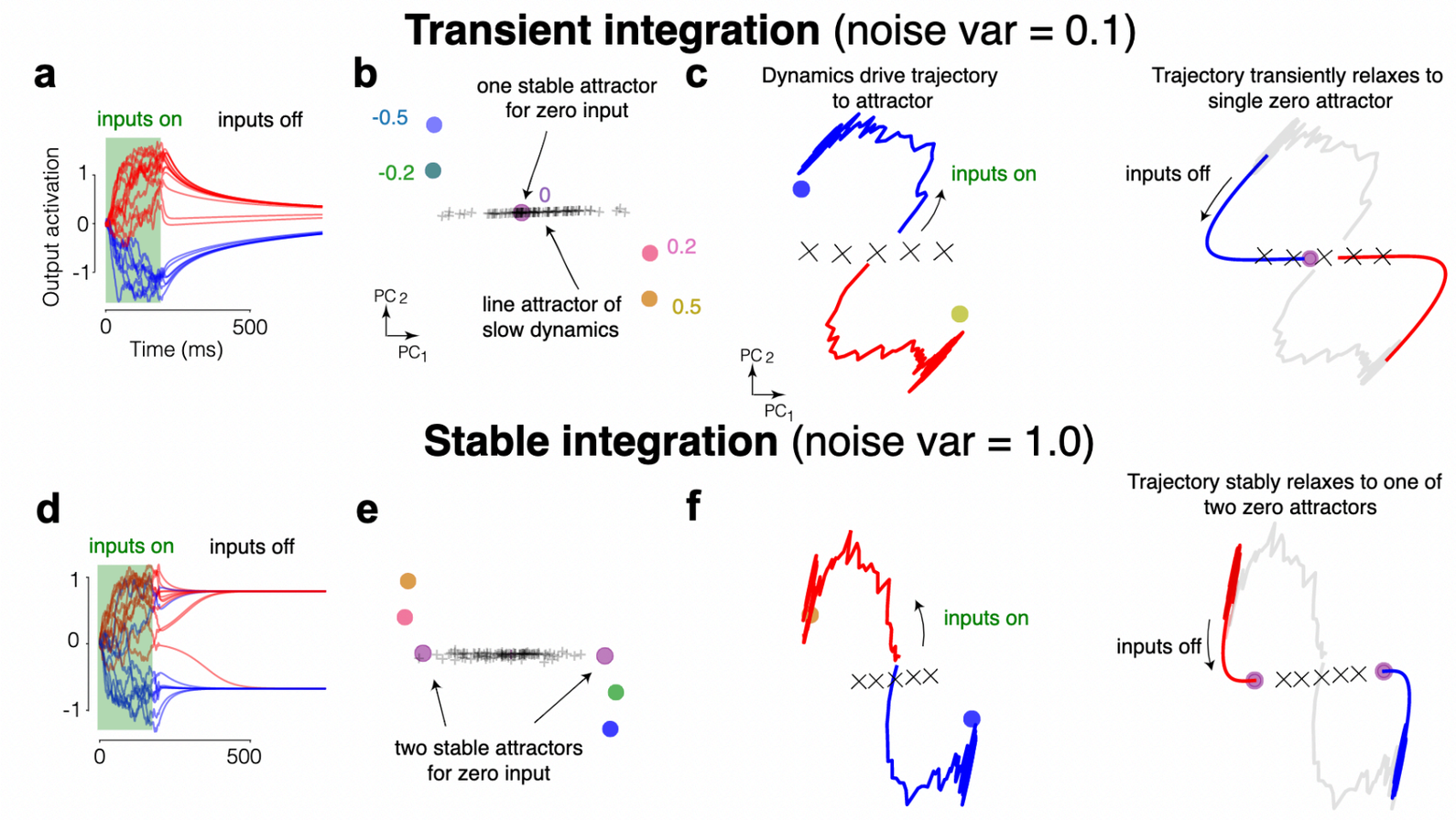
Learning rule



McMahan, Kleinman, Kao, NeurIPS 2021

Another related observation we won't go into detail about

An observation: input noise affects the integration mechanism.
For more detail, see McMahan et al, NeurIPS 2021.

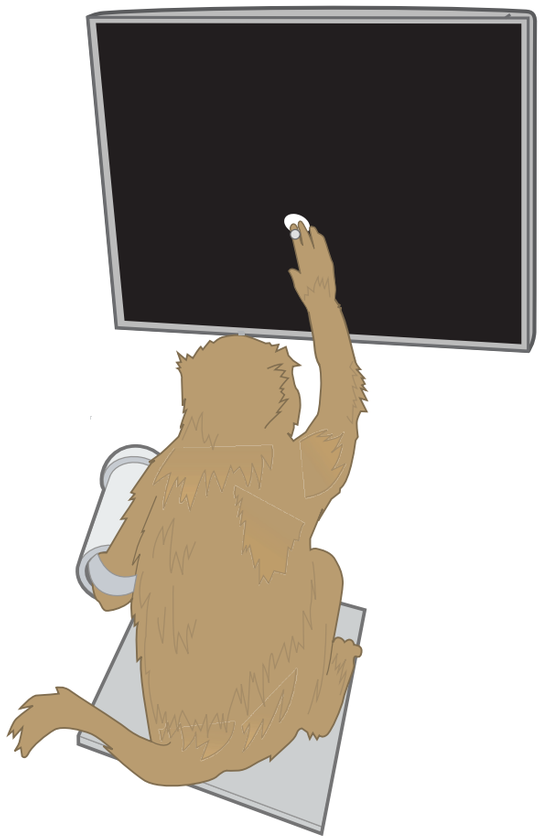


McMahan, Kleinman, Kao, NeurIPS 2021

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*

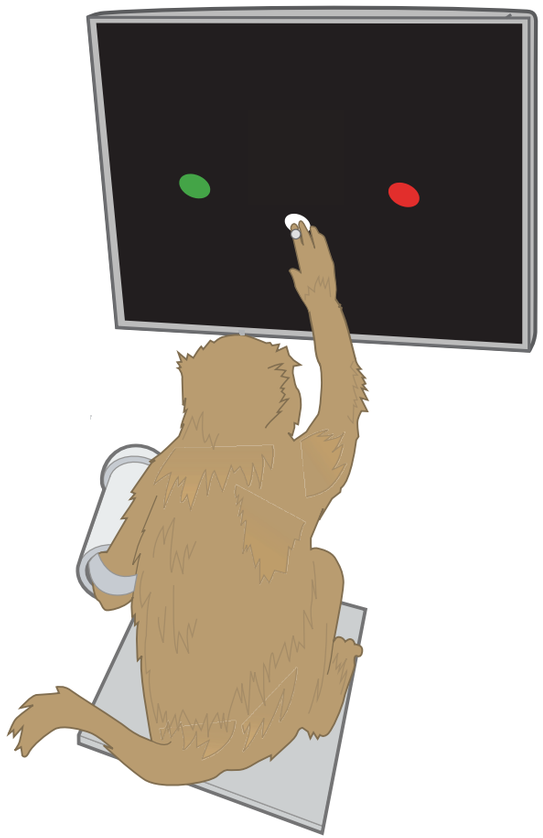
Using RNNs to study multi-area computation.

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*



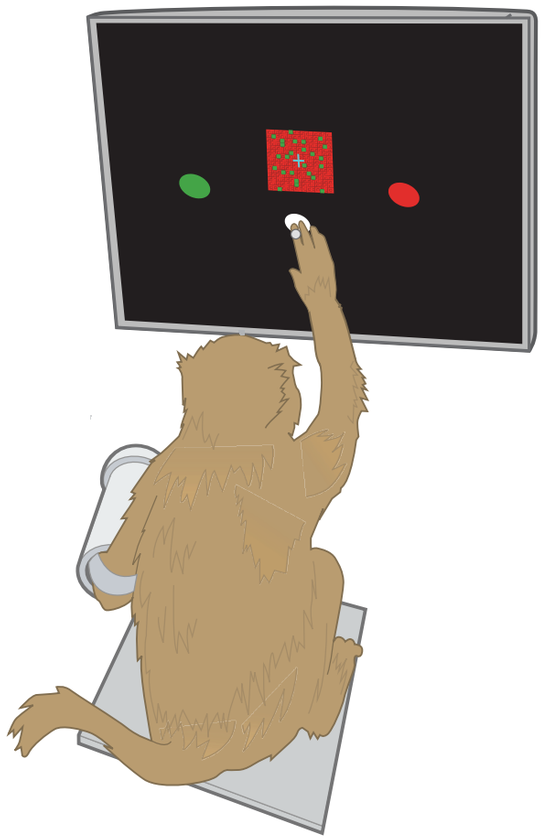
Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*



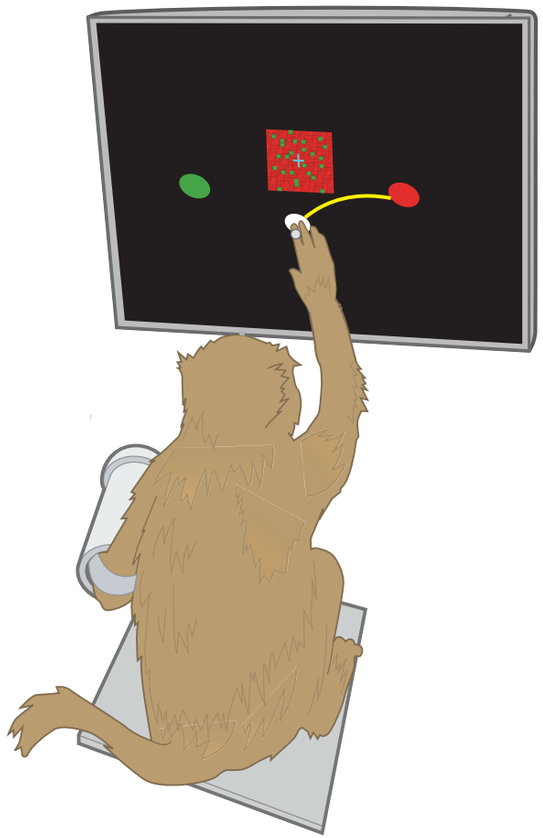
Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*



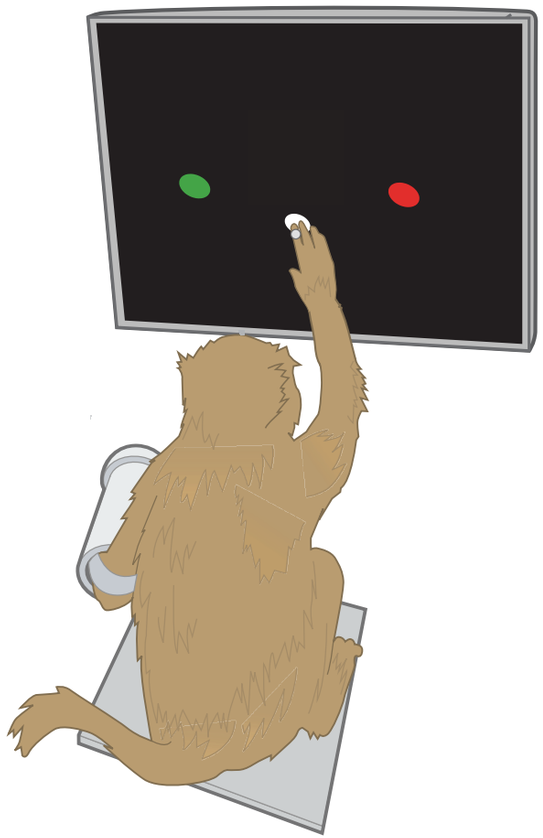
Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*



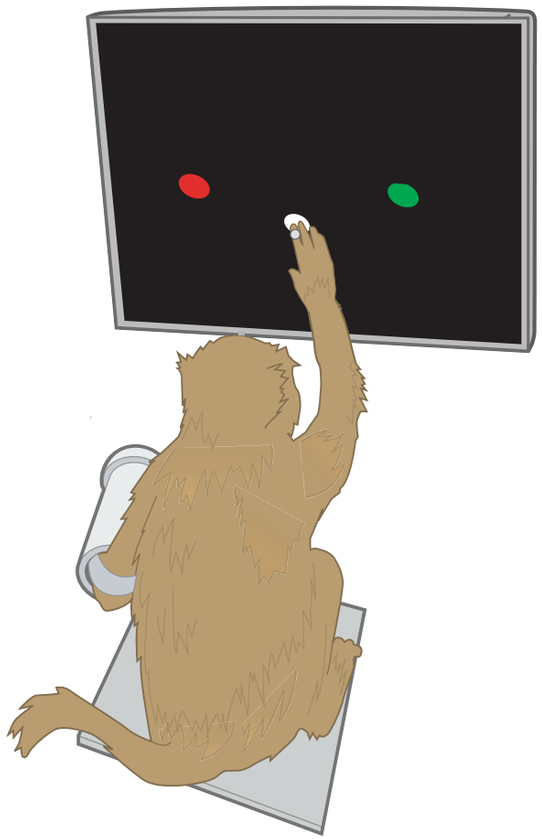
Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*



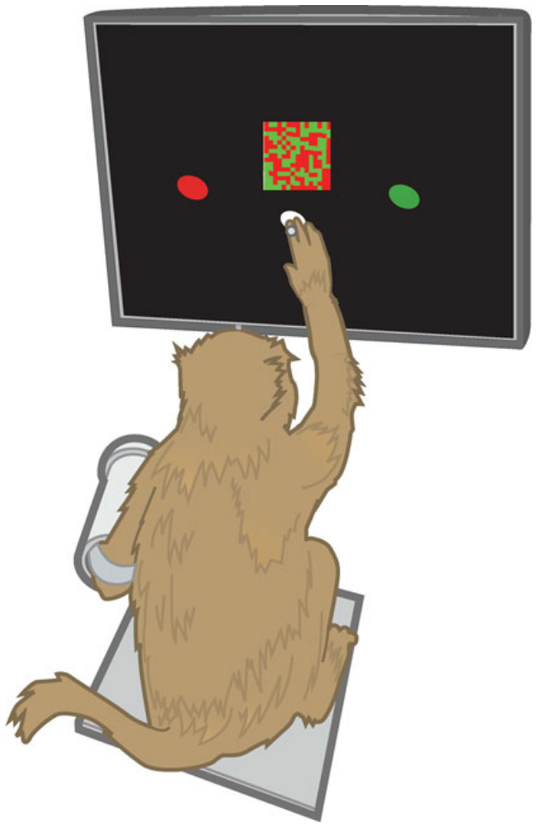
Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*



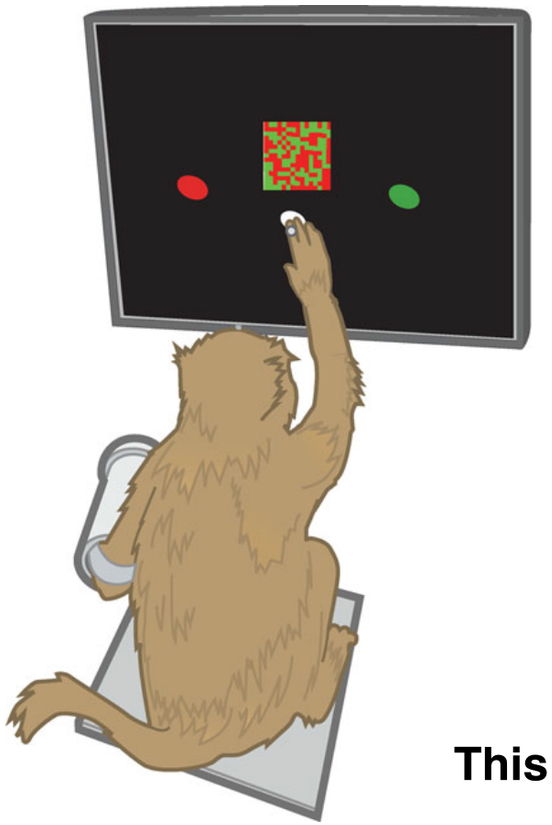
Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

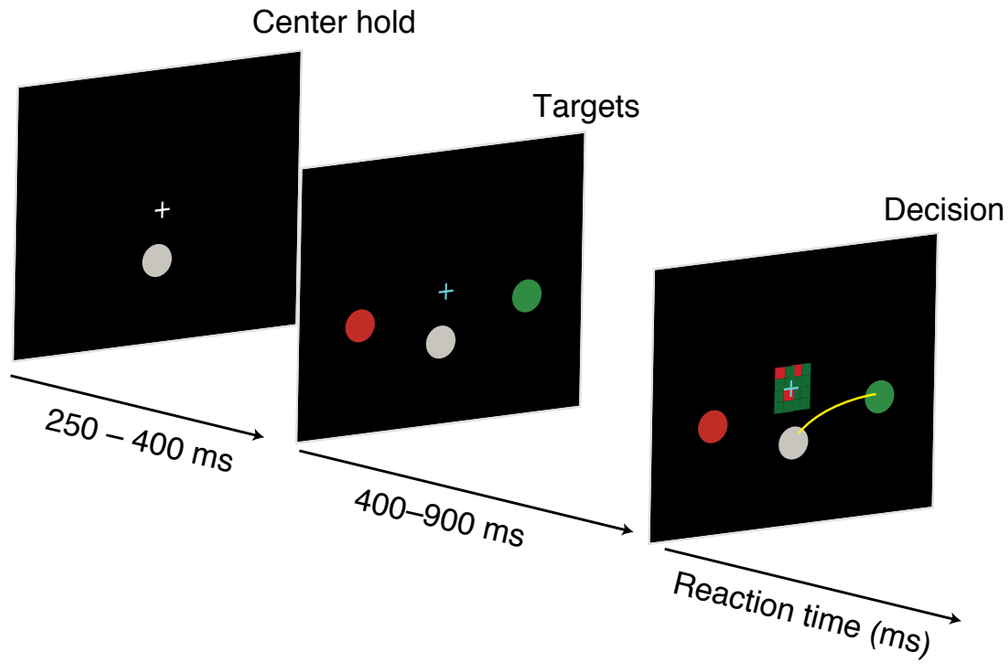
Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*



Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*

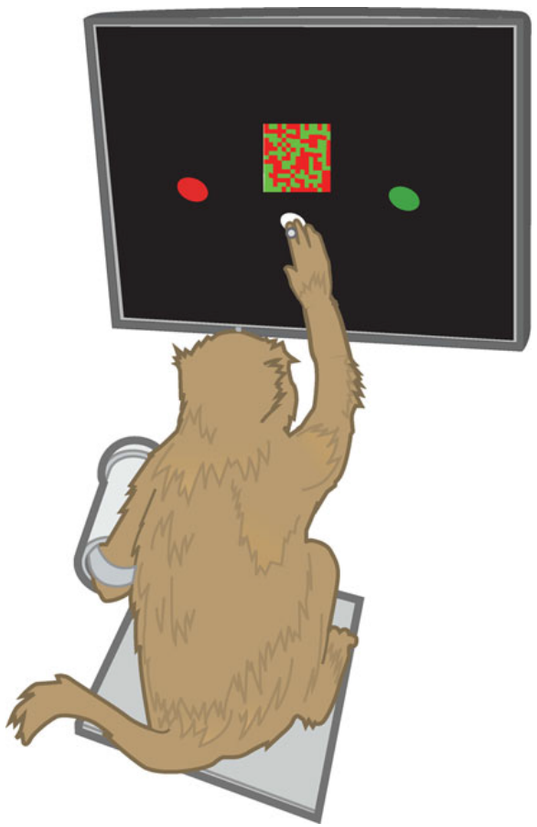


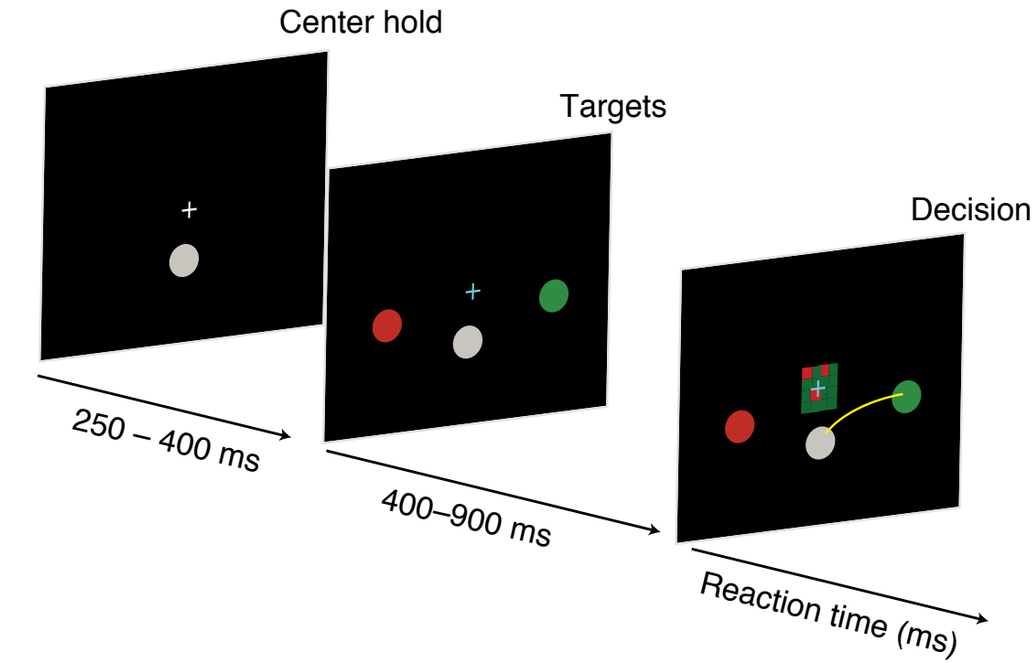


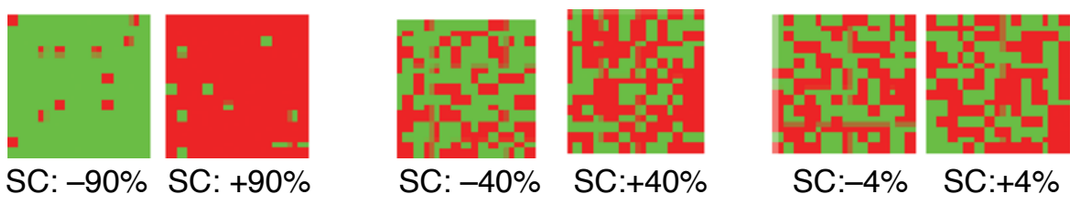
This task dissociates the color choice and the motor choice.

Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*

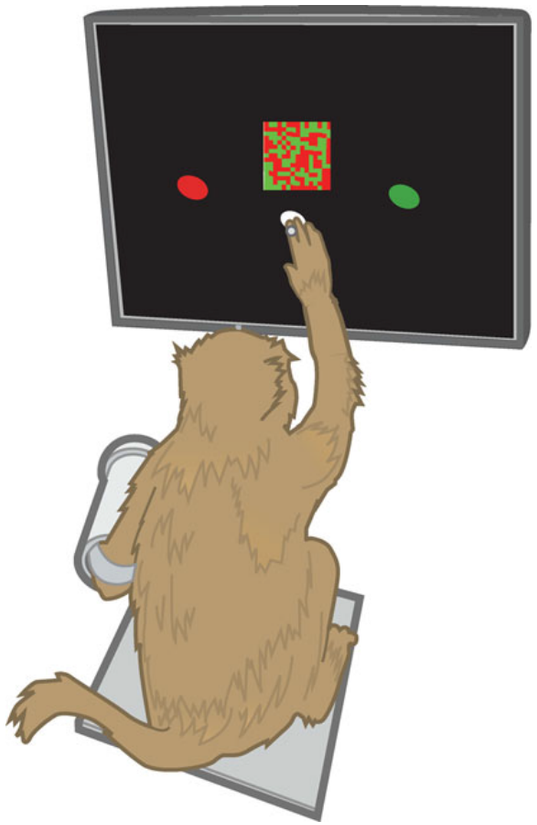


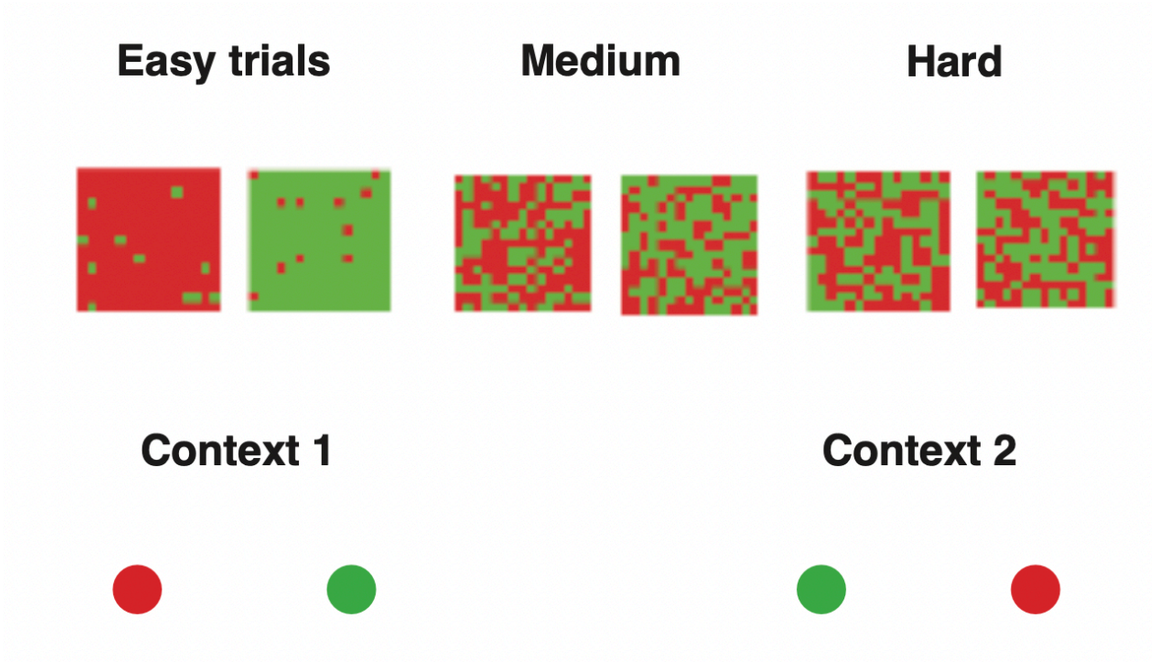




Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

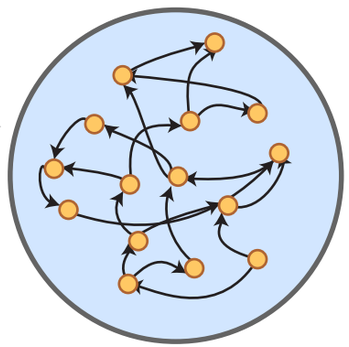
Example 2: Kleinman, Chandrasekaran*, Kao*, *NeurIPS 2021*



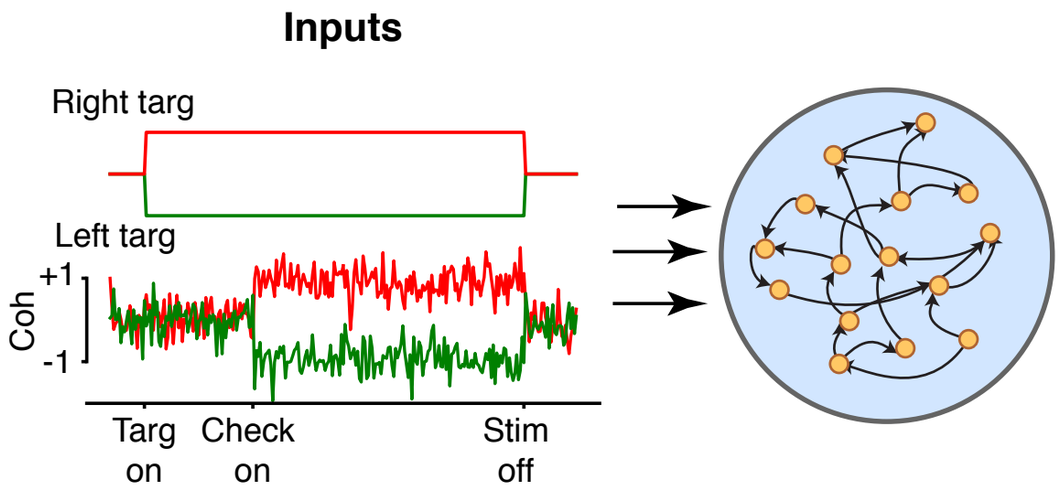


Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

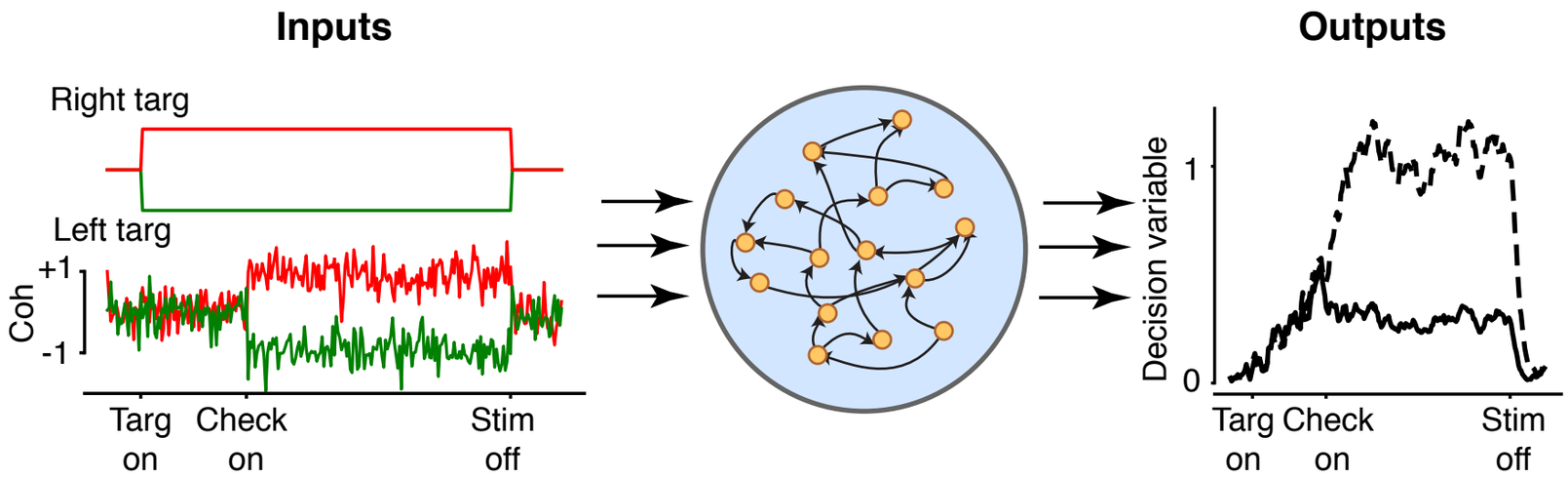
Let’s train an RNN for this task

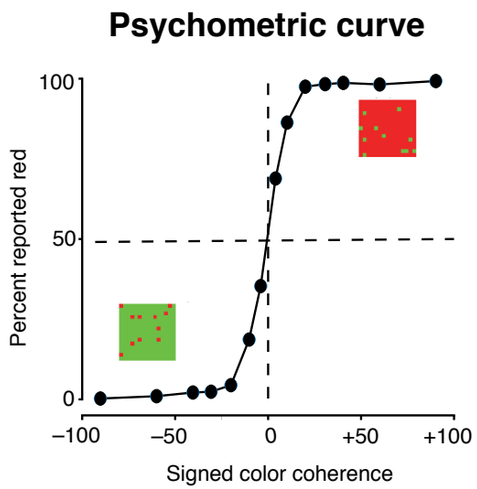


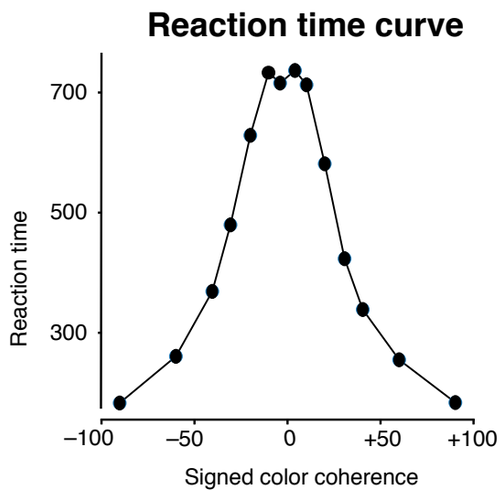
Let’s train an RNN for this task



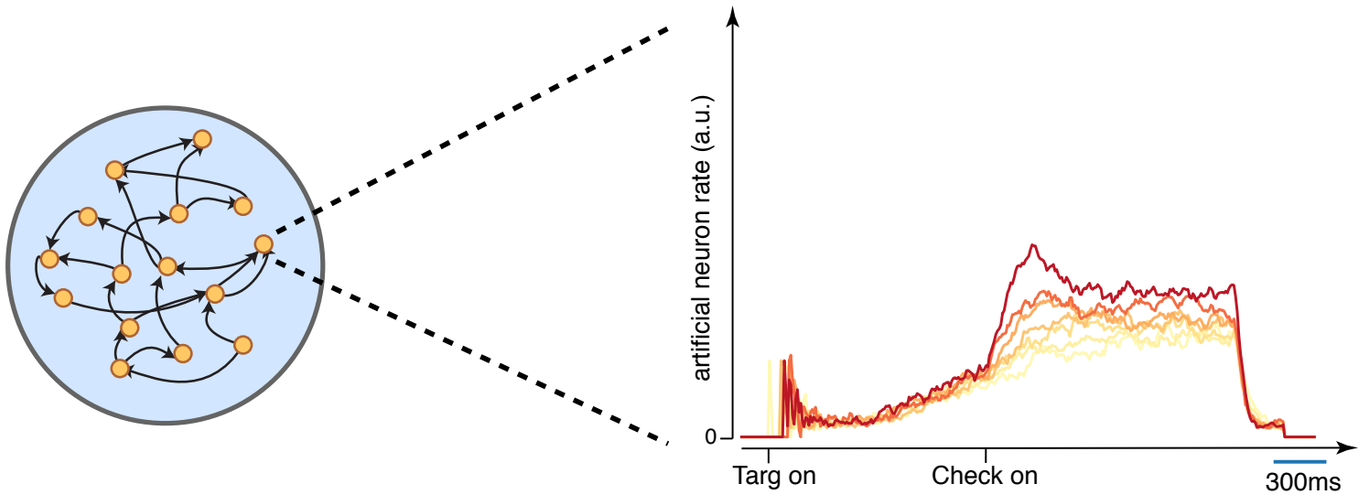
Let's train an RNN for this task



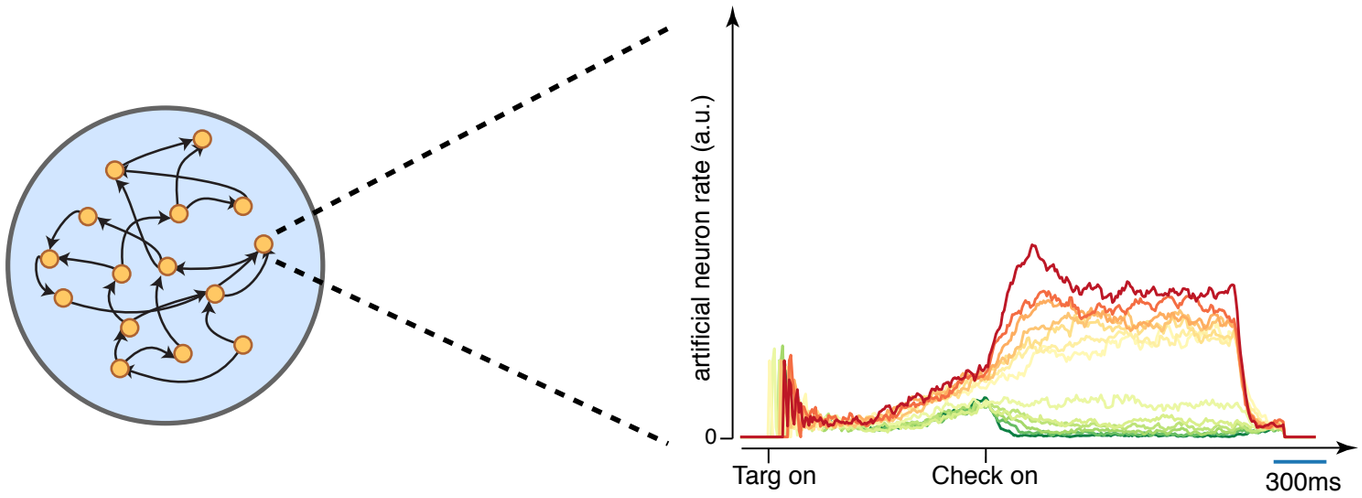




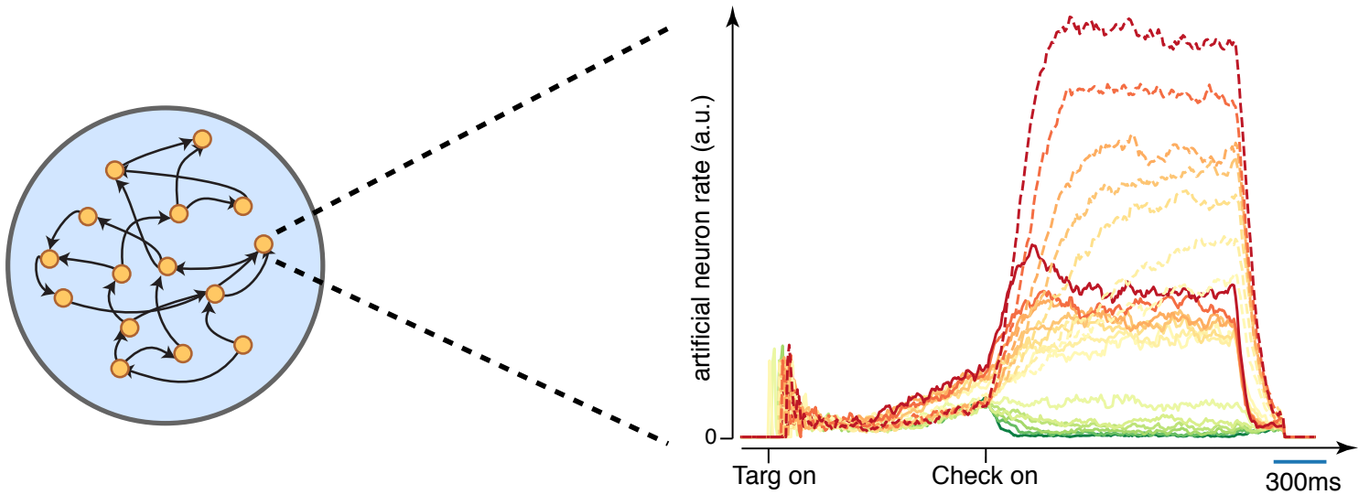
Artificial neuron activity



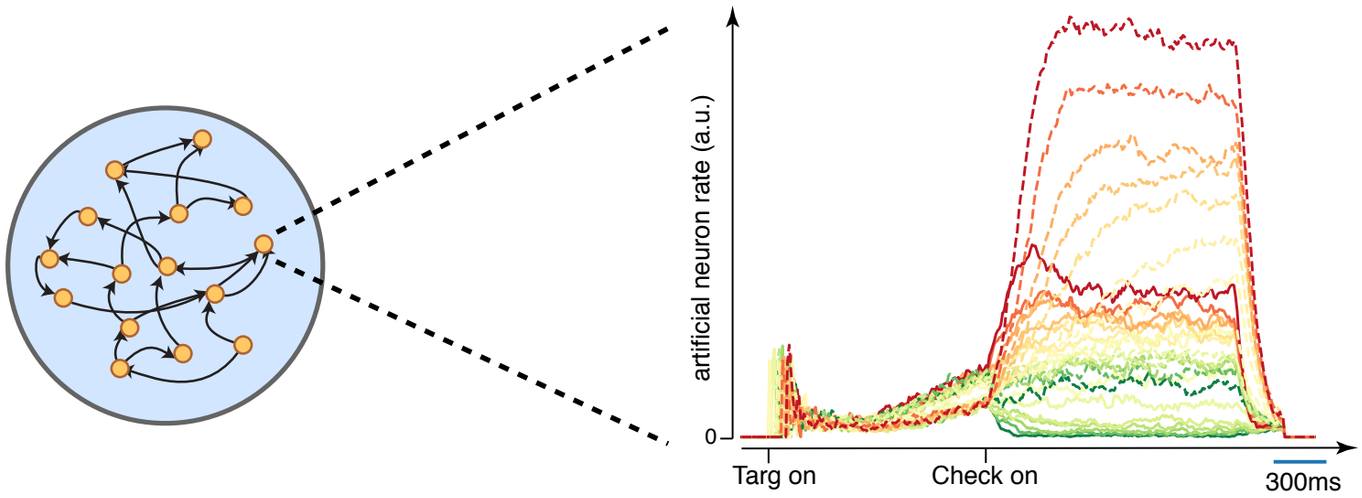
Artificial neuron activity



Artificial neuron activity



Artificial neuron activity

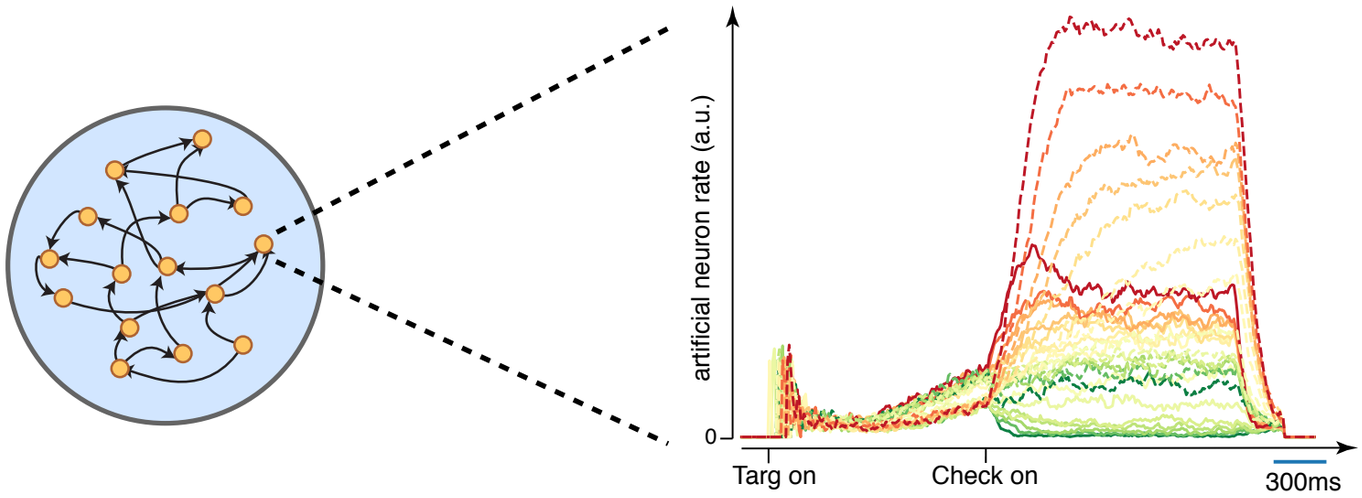


This artificial unit is primarily **color selective**.

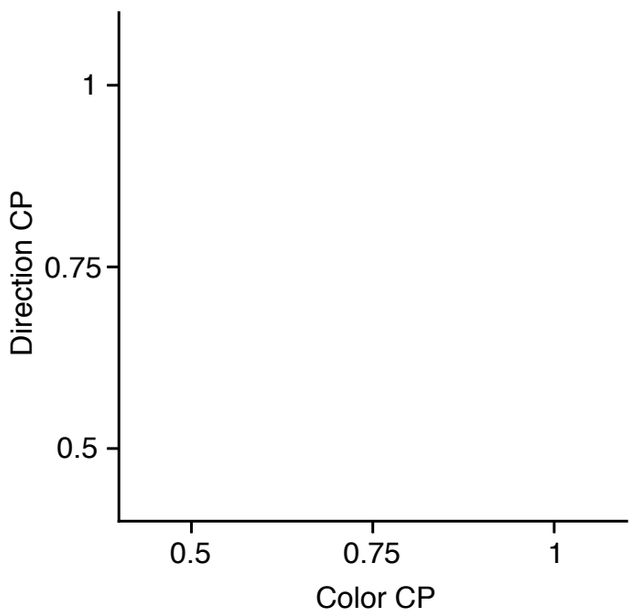
Color choice probability (CP): 0.95

Direction CP: 0.70

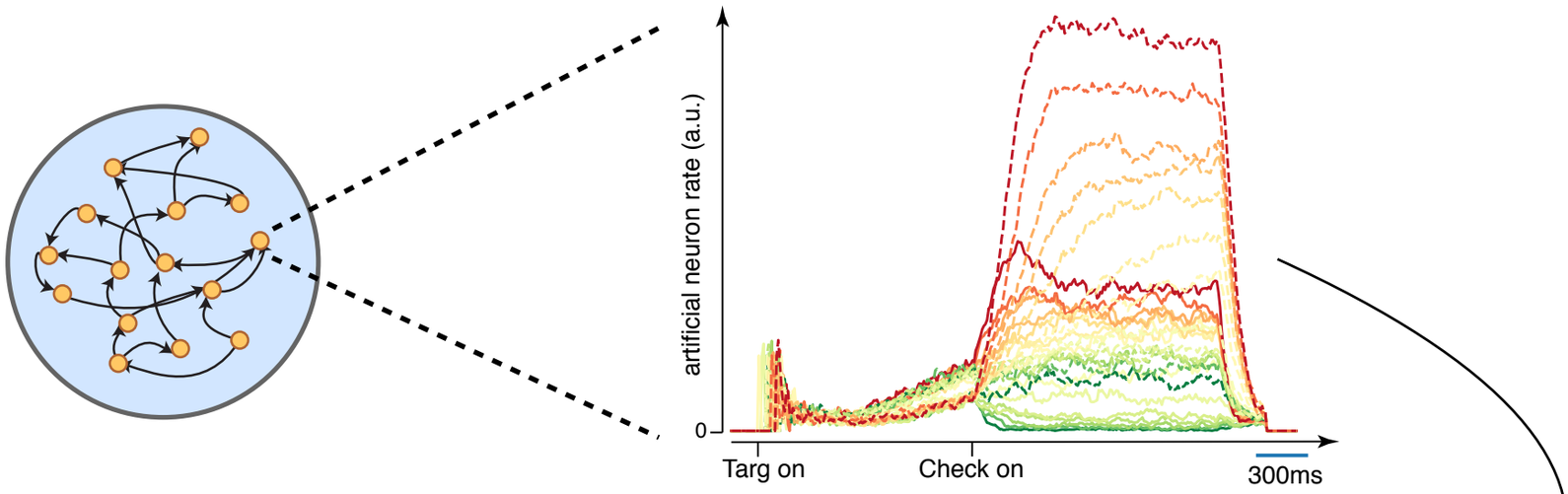
Artificial neuron activity



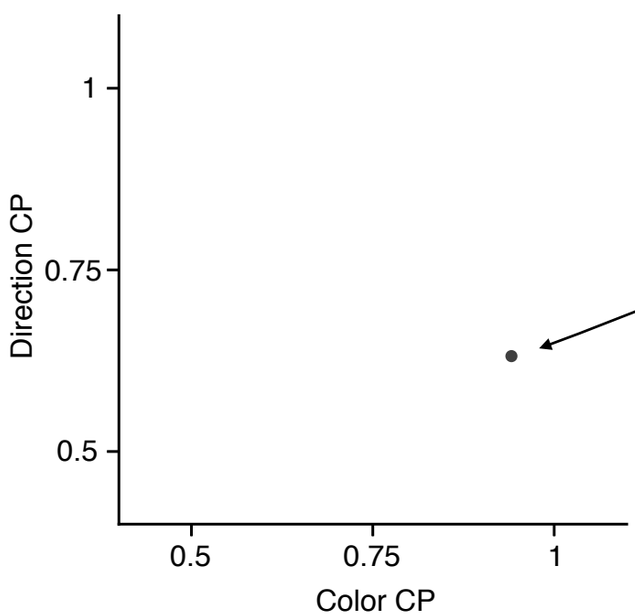
Choice probabilities



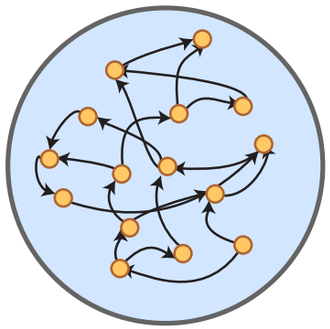
Artificial neuron activity

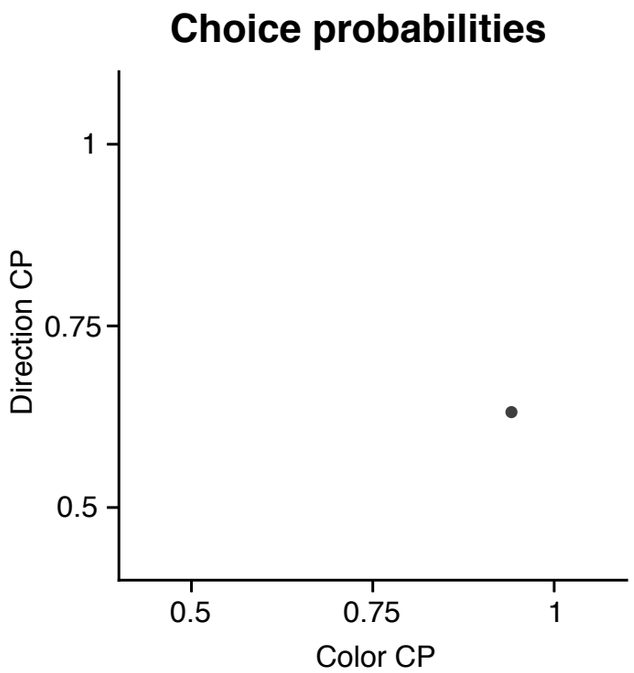


Choice probabilities

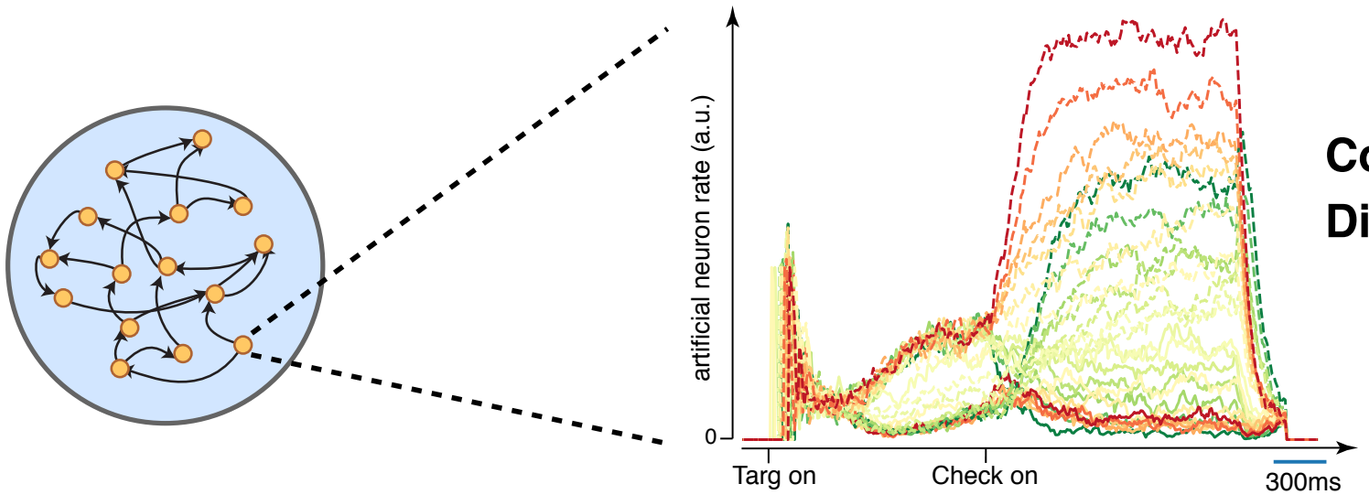


Artificial neuron activity





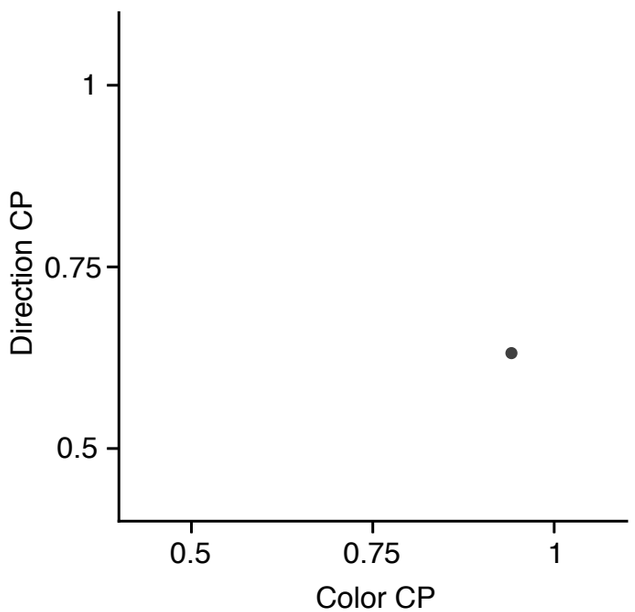
Artificial neuron activity



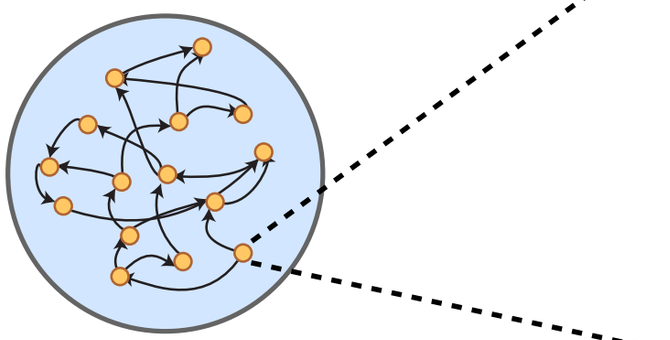
Color CP: 0.56

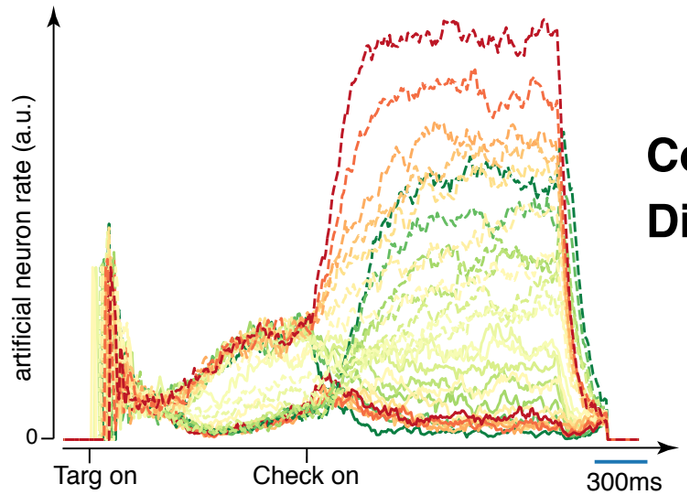
Direction CP: 0.88

Choice probabilities



Artificial neuron activity

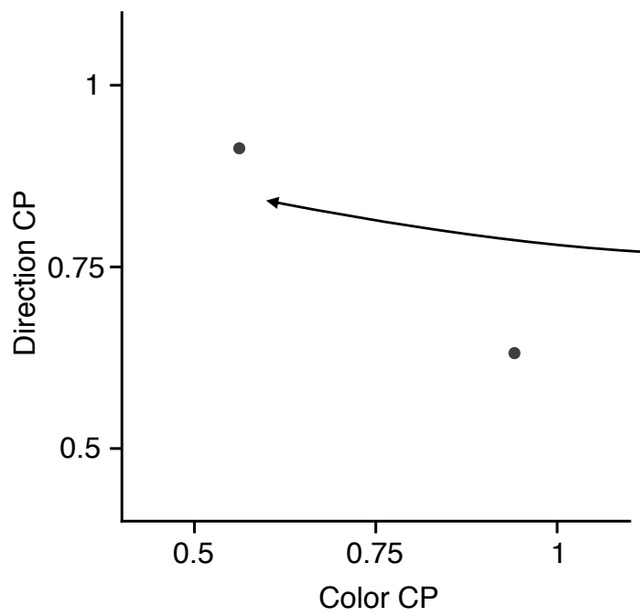




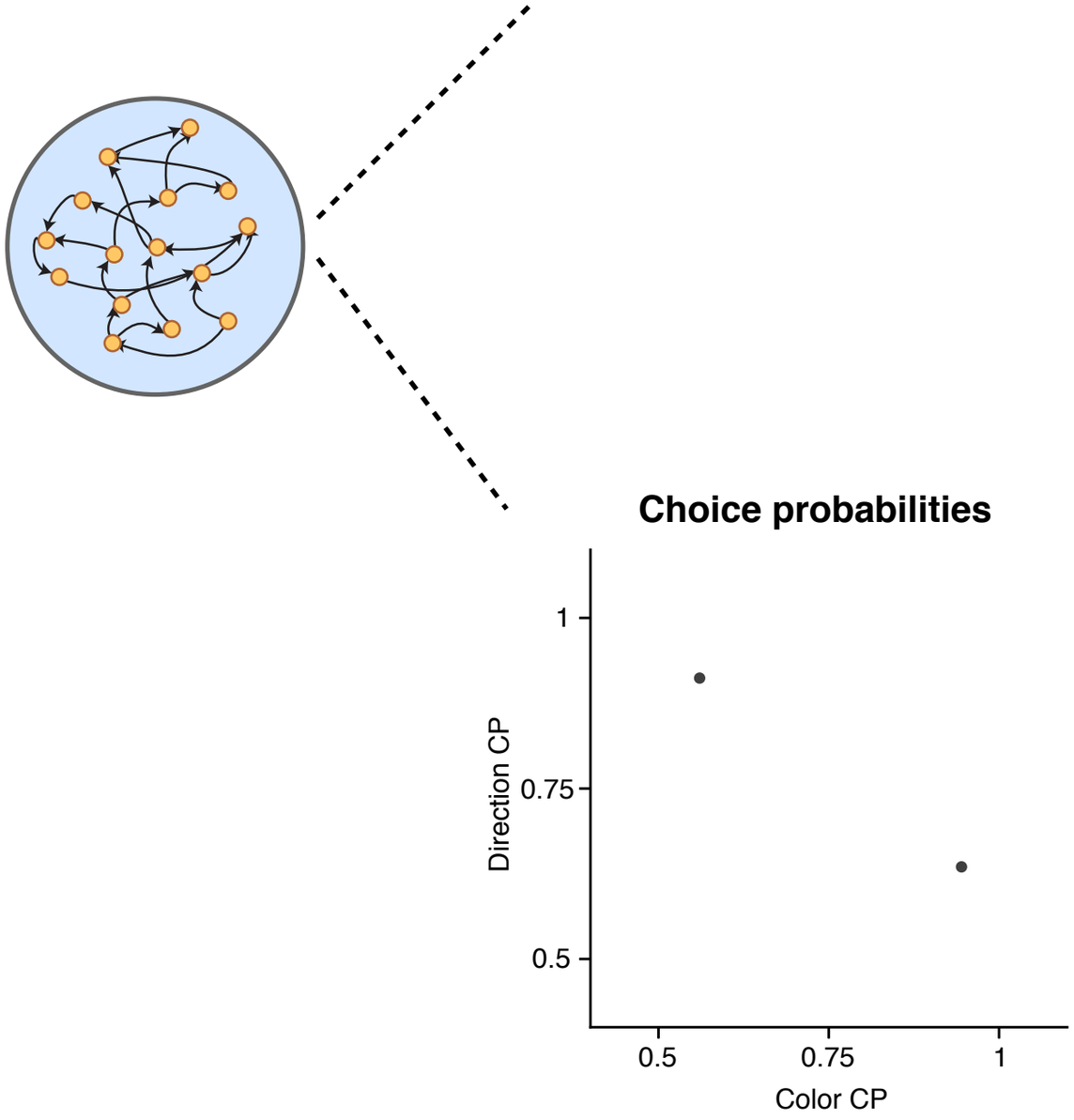
Color CP: 0.56

Direction CP: 0.88

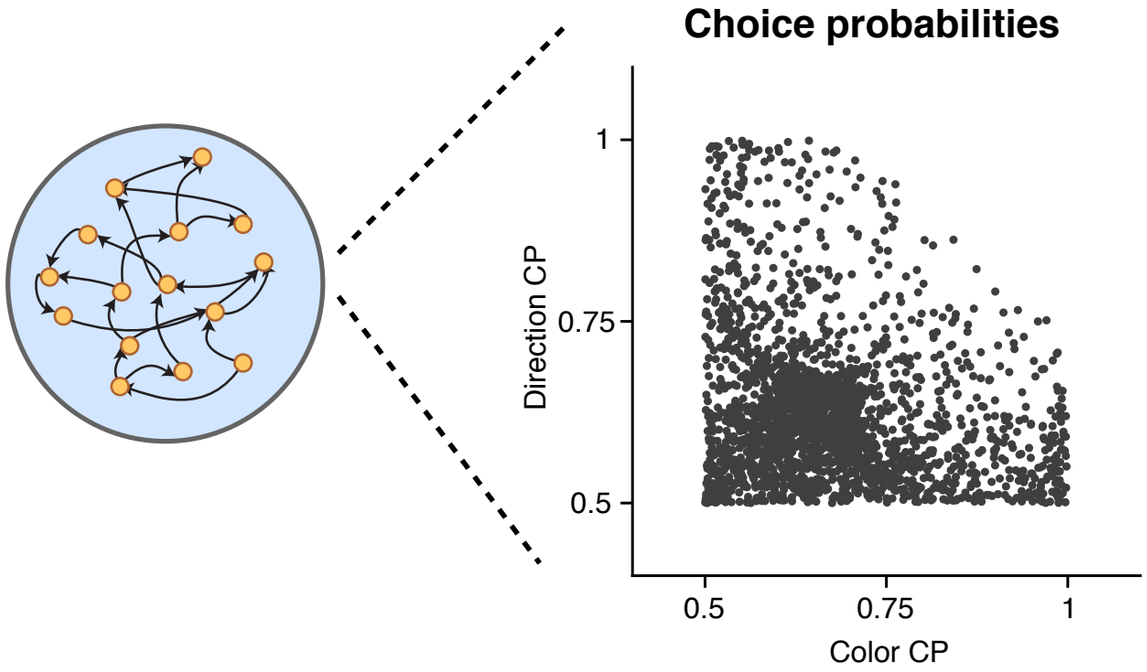
Choice probabilities



Artificial neuron activity

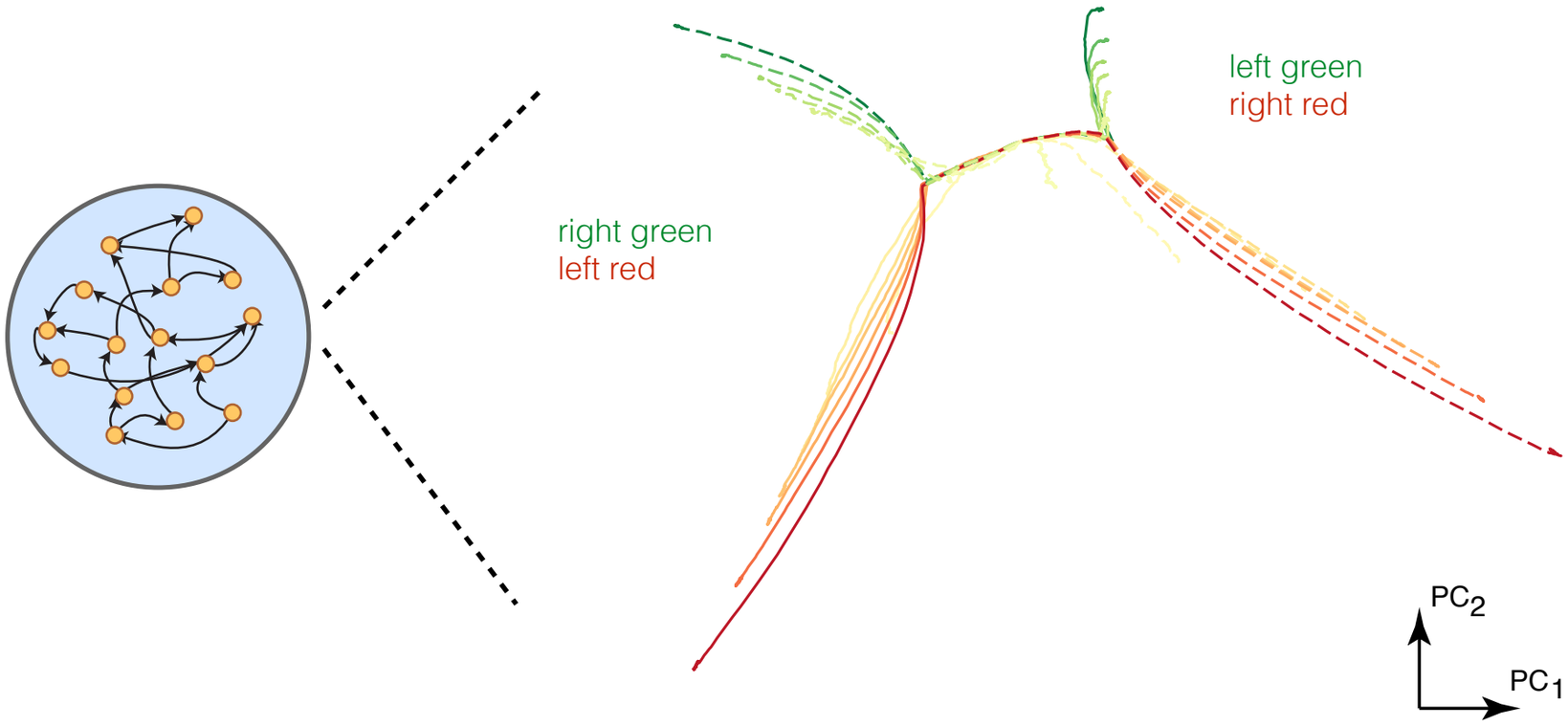


Artificial neuron activity

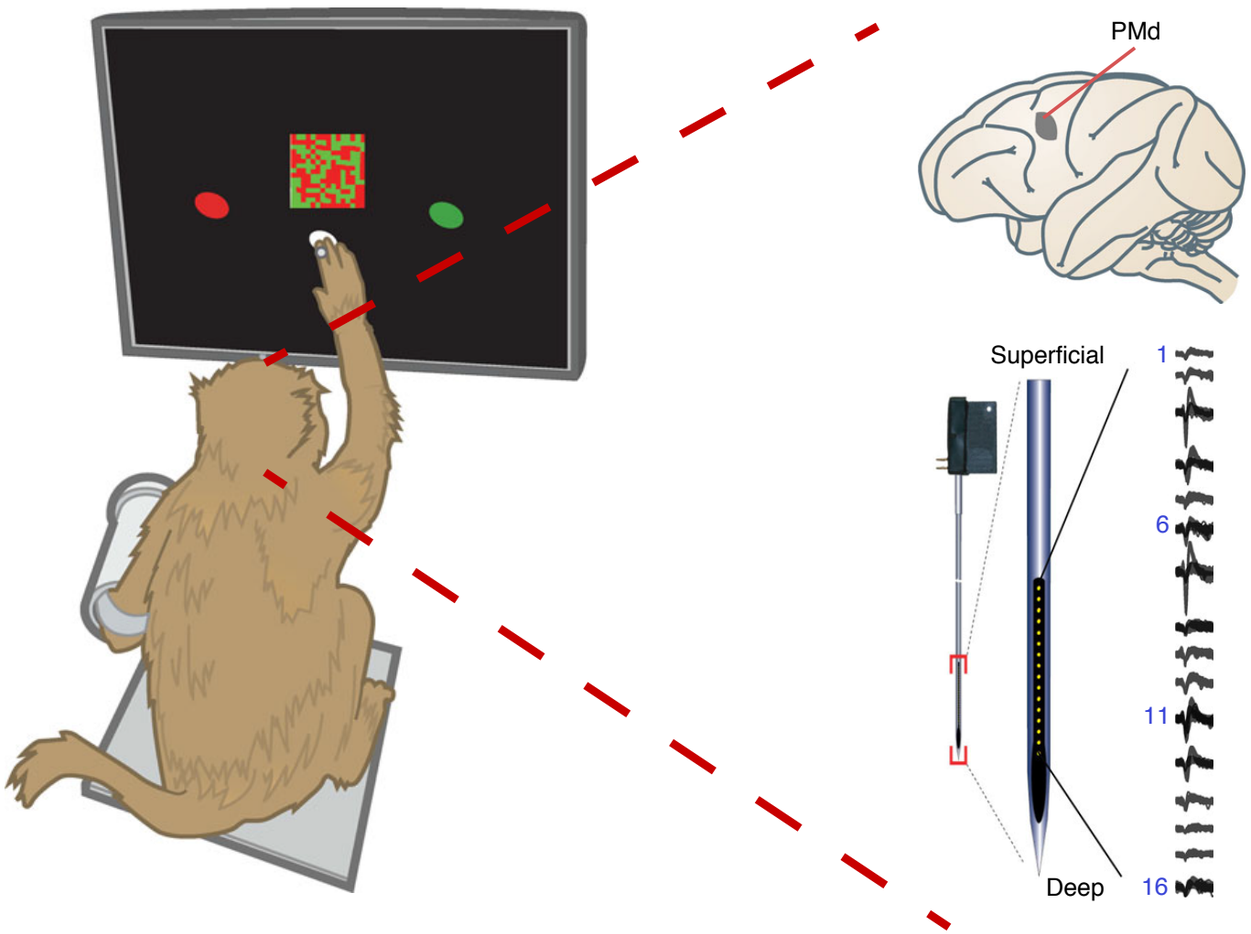


This RNN uses artificial units with a mix of color and direction selectivity.

Artificial neuron activity

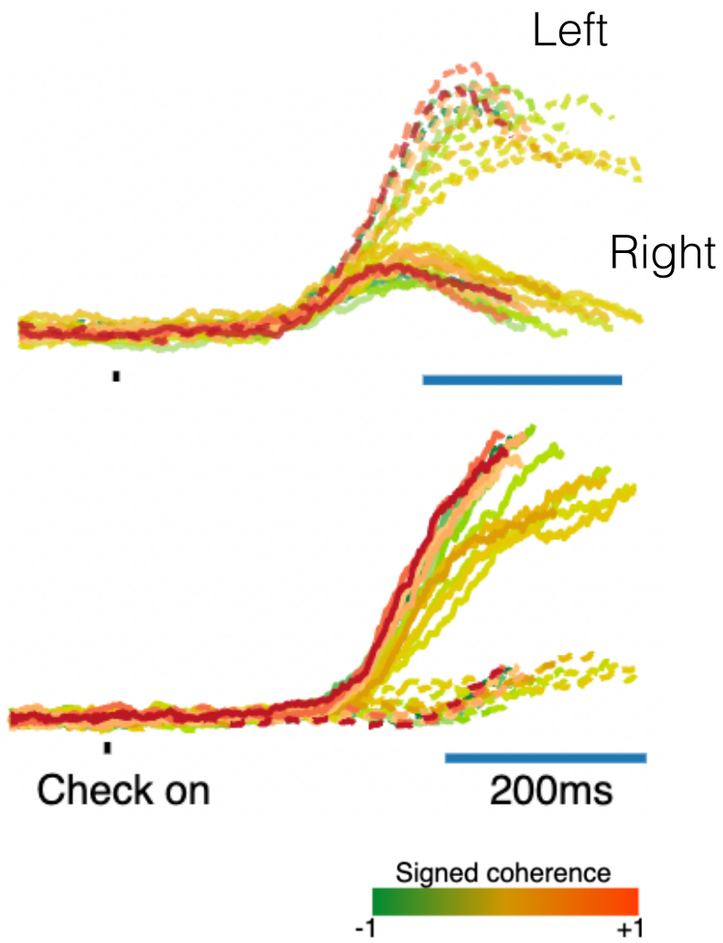


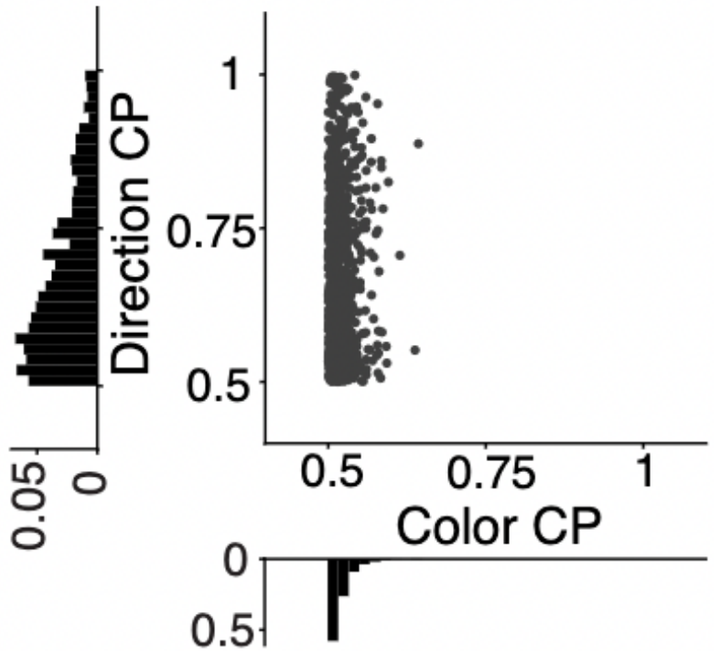
PMd recordings



Coallier and Kalaska, 2015; Chandrasekaran, Peixoto, Newsome, Shenoy, Nature Communications, 2017

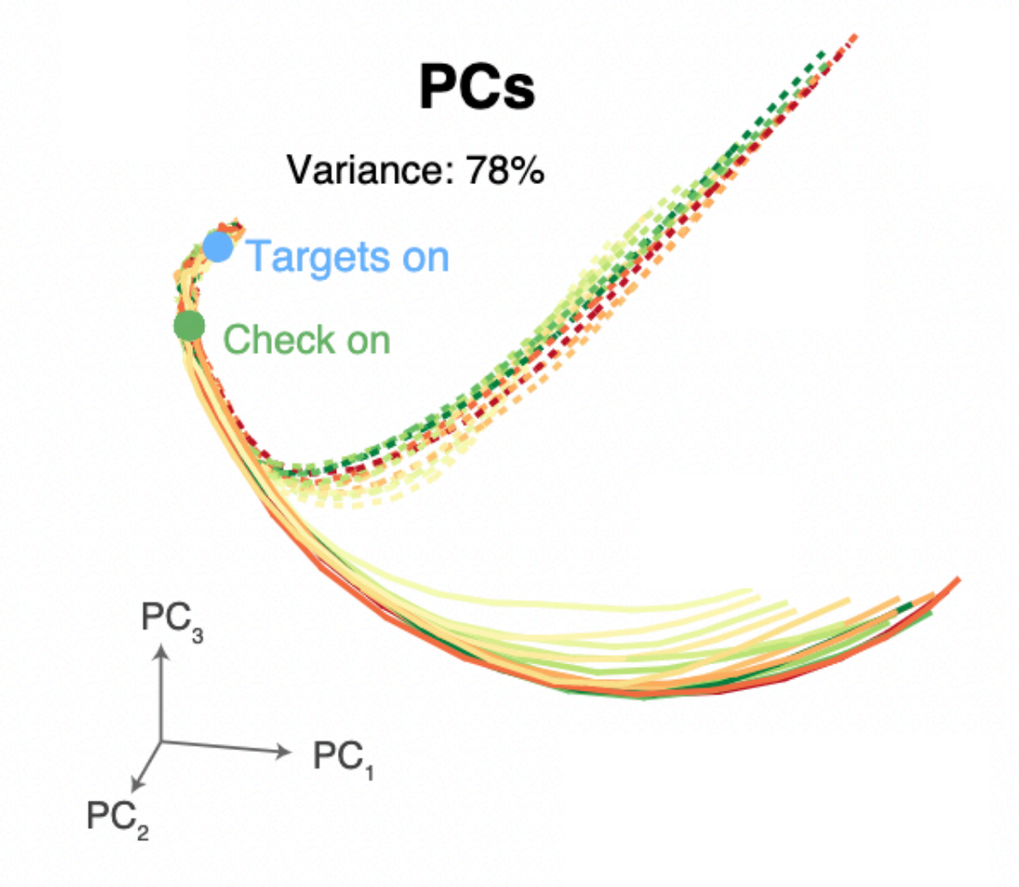
PMd recordings

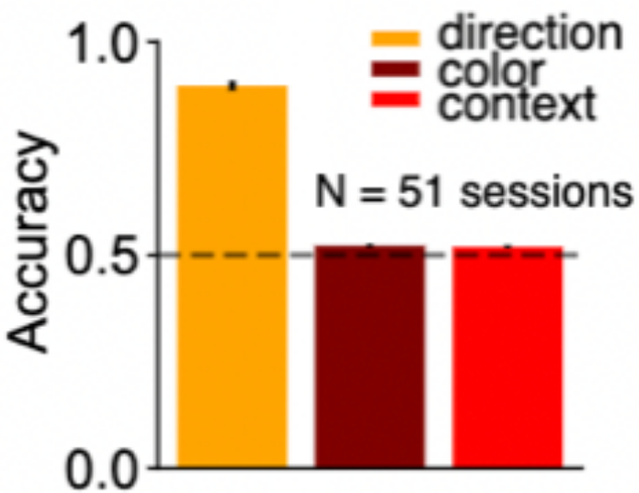




Single neuron choice probabilites are near chance for the color decision.

PMd recordings

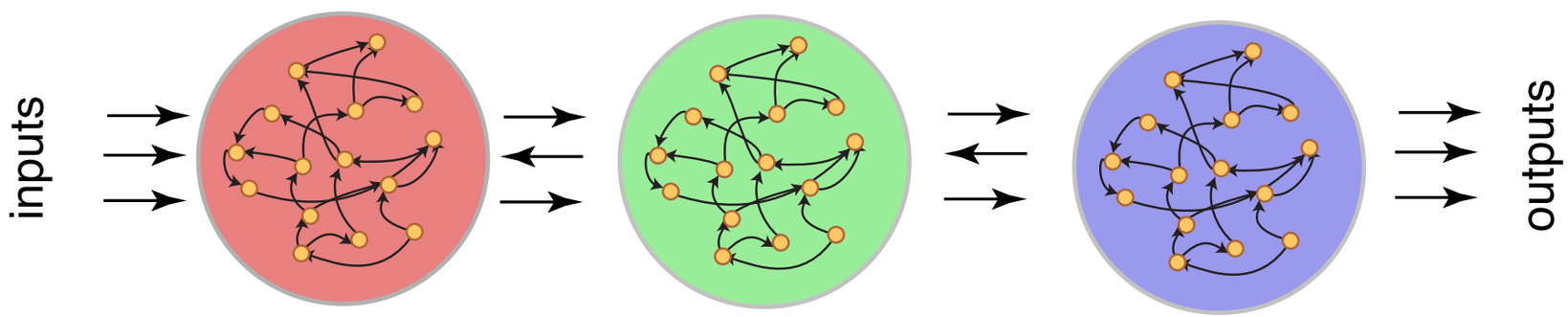




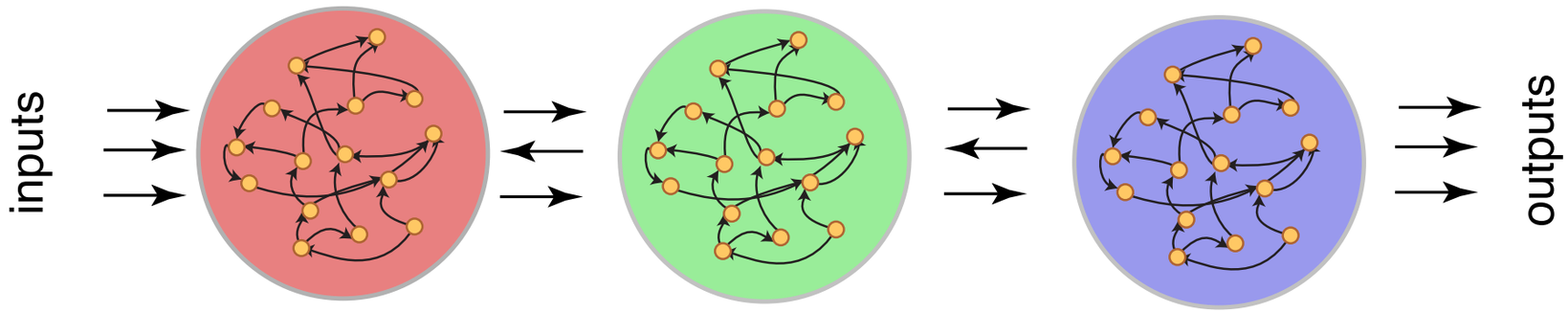
PMd exhibits two trajectory motifs, related to the direction decision. Accuracy for decoding context and color are not significantly above chance.

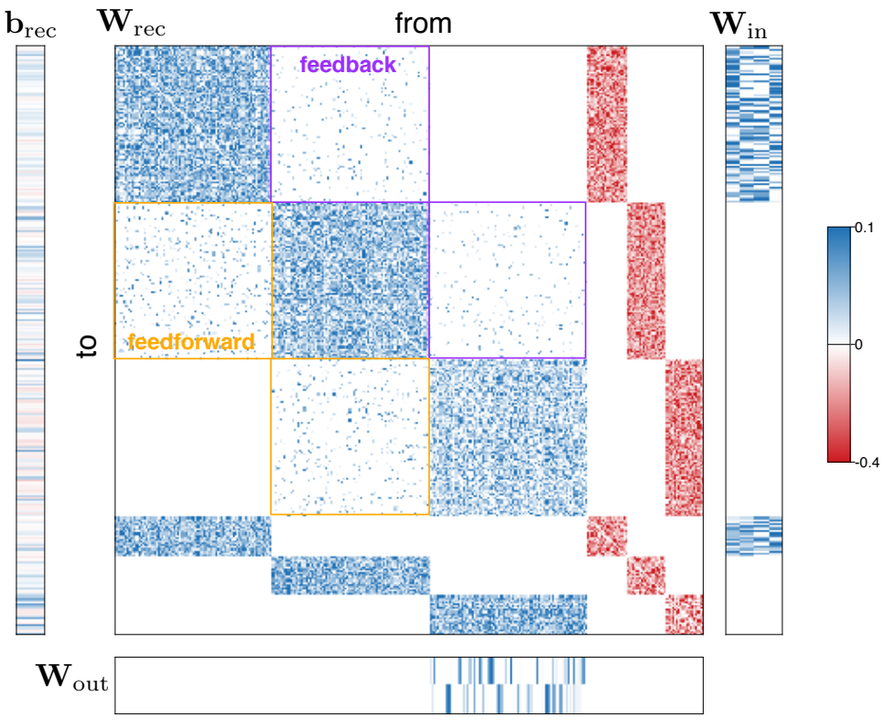
In this case, artificial neurons do not resemble electrophysiology

Idea: color information is required to solve the task, so it must be used in an upstream area. What if we incorporate multiple areas in the RNN?

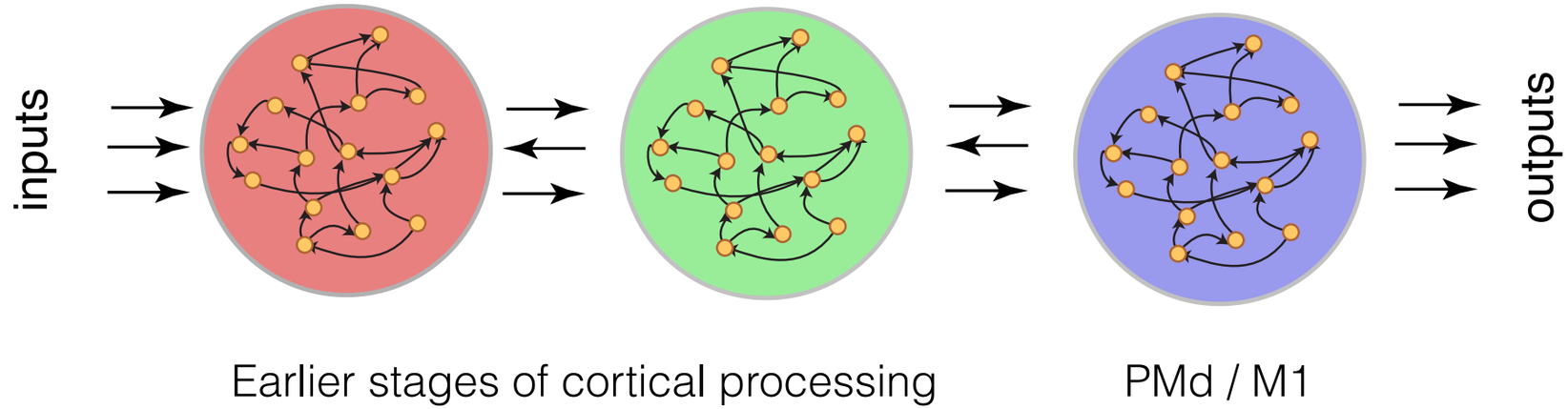


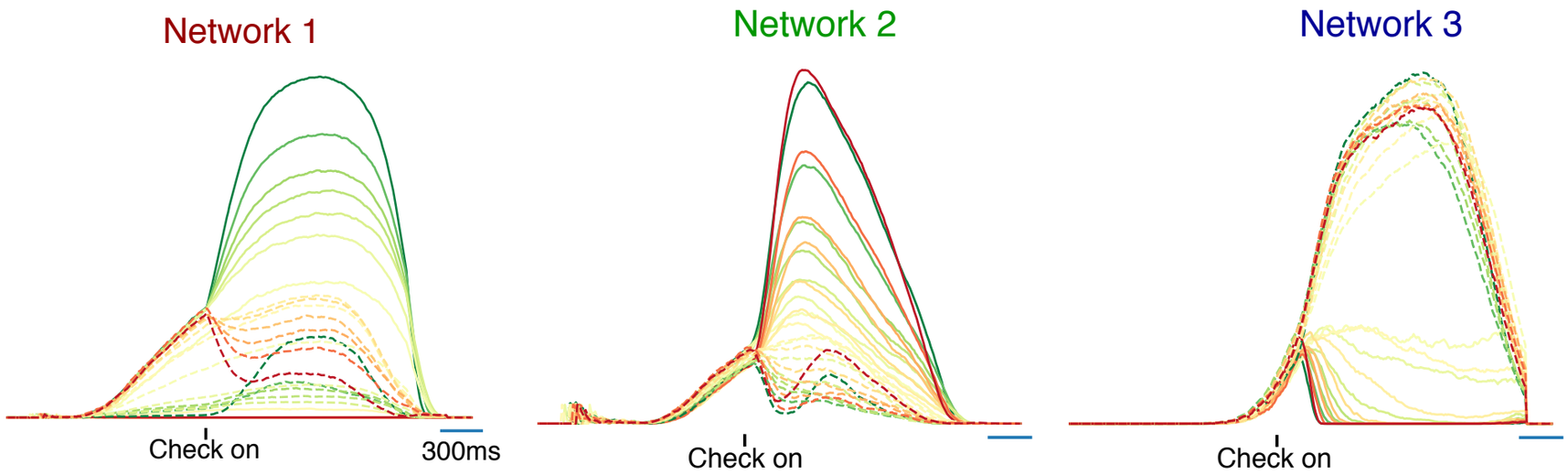
In this case, artificial neurons do not resemble electrophysiology



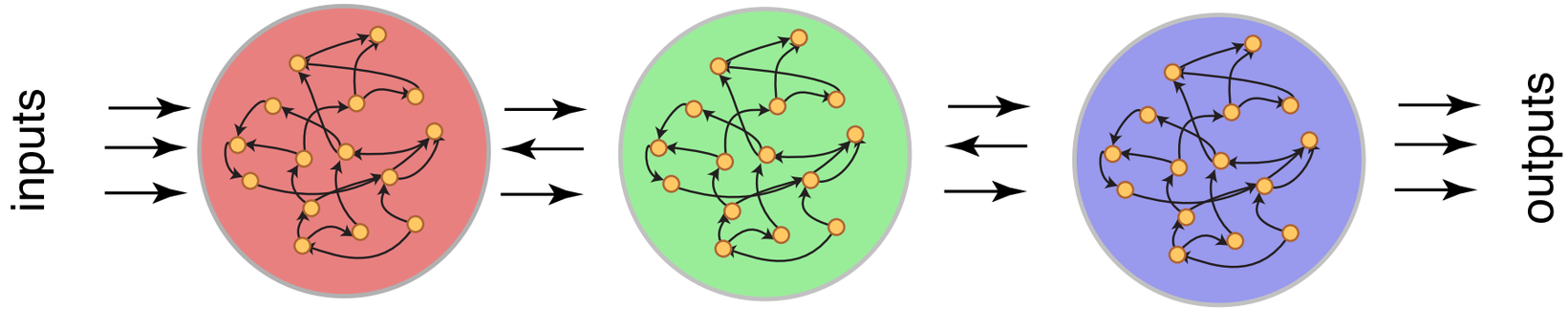


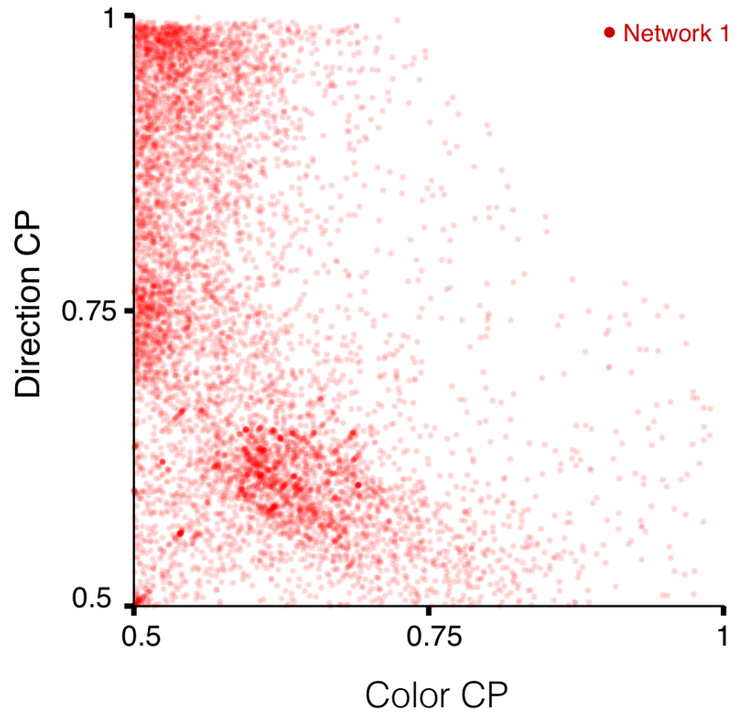
Using multiple areas lead to PMd resembling responses



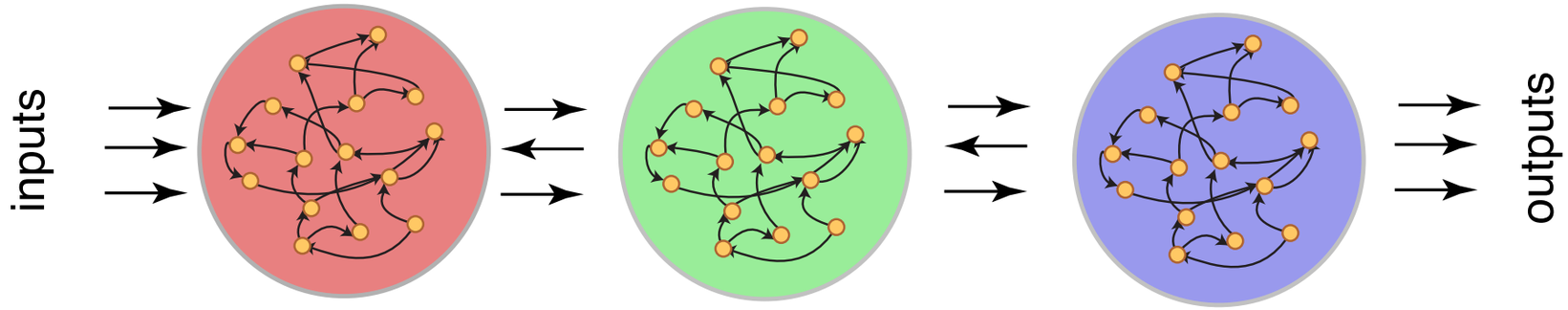


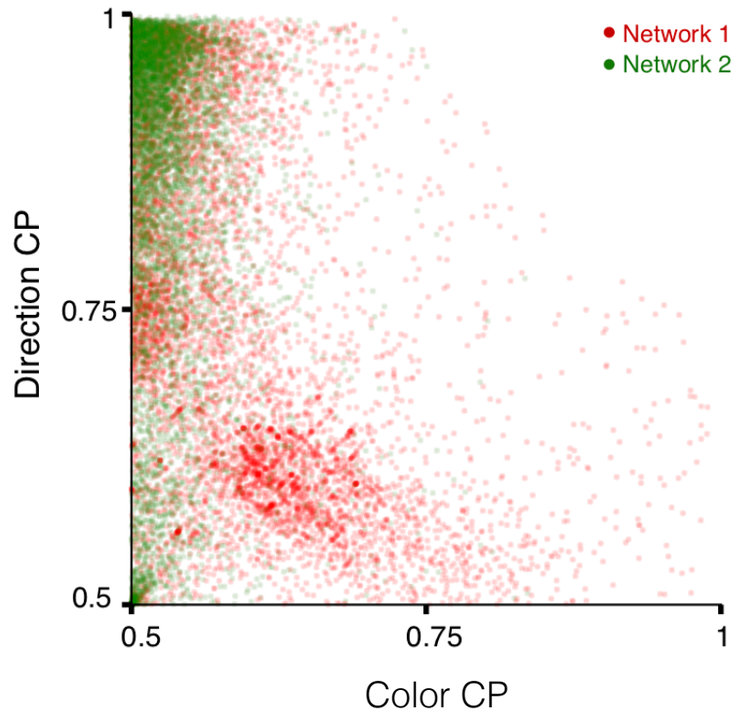
Using multiple areas lead to PMd resembling responses



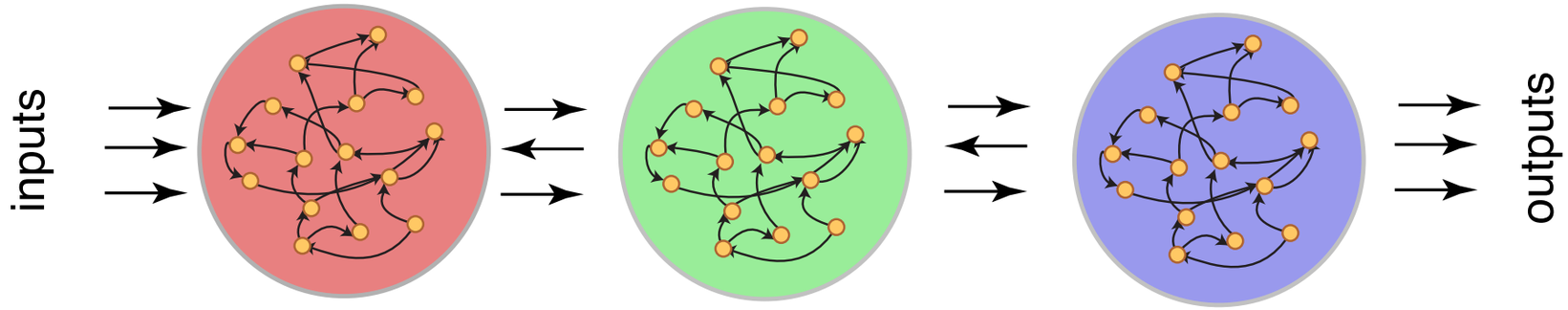


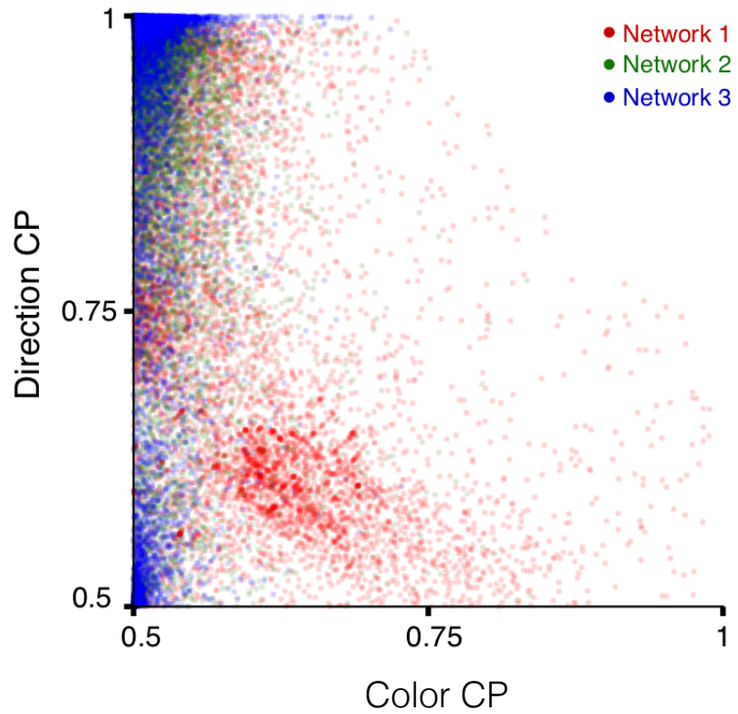
Using multiple areas lead to PMd resembling responses



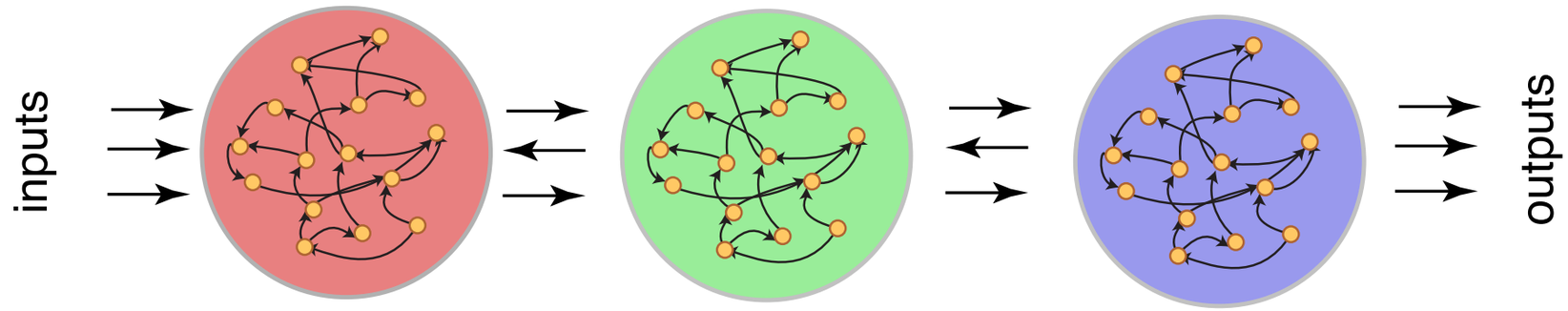


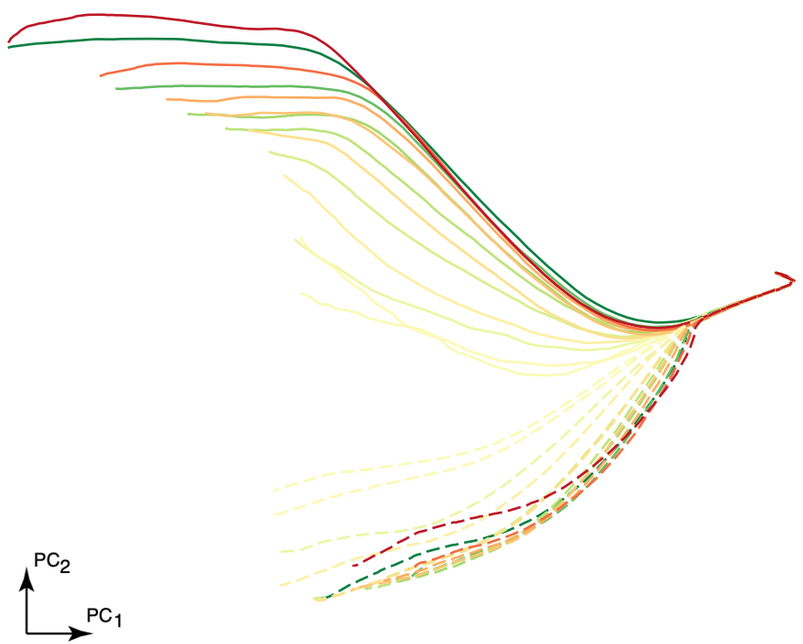
Using multiple areas lead to PMd resembling responses



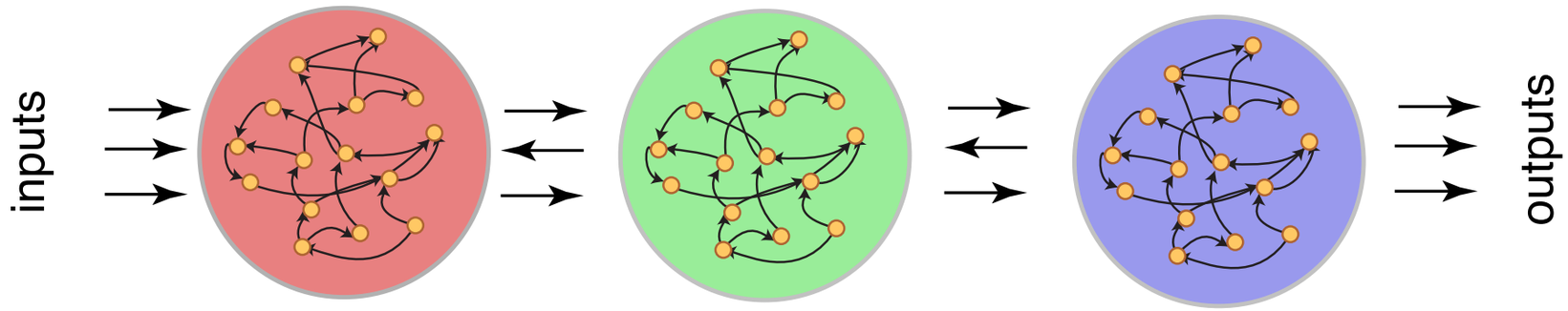


Using multiple areas lead to PMd resembling responses





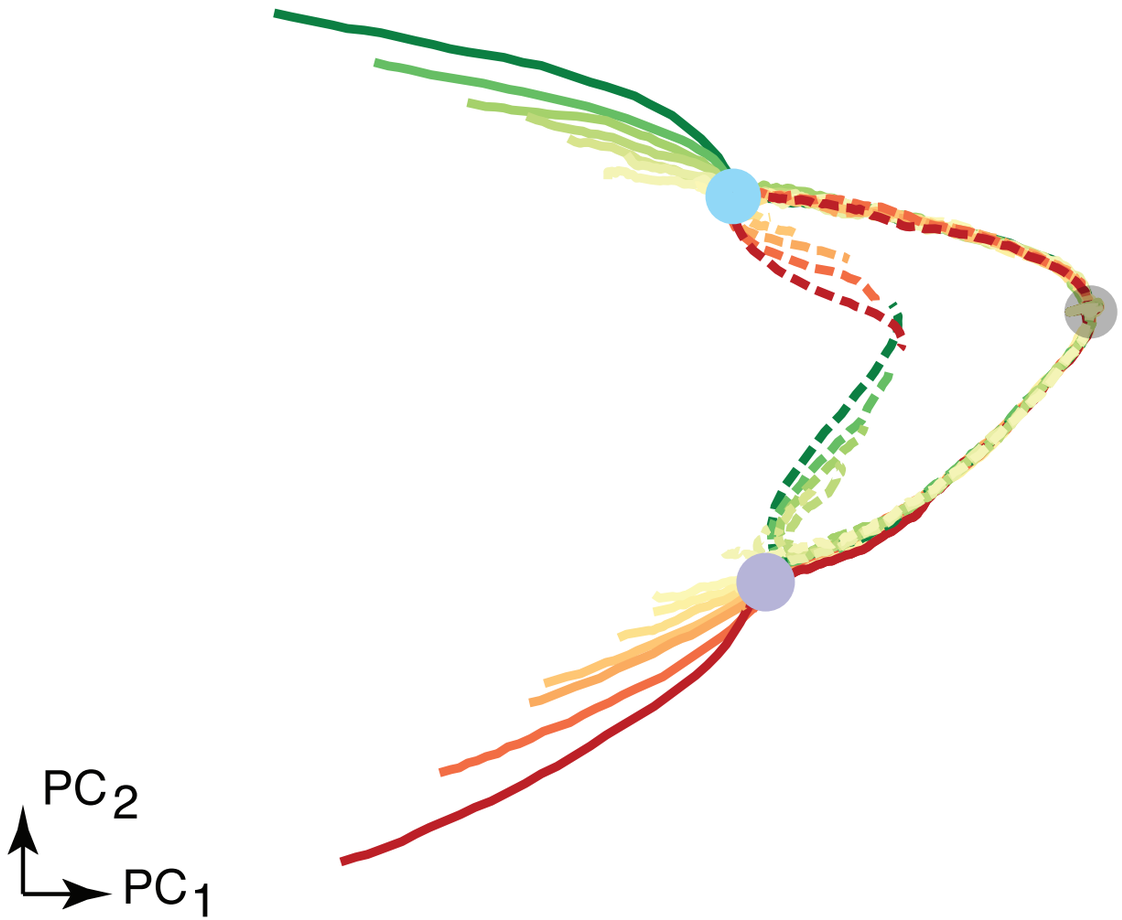
What about the computation to turn color into direction?



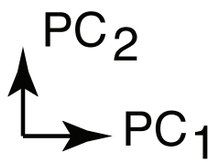
What mechanism achieves this color filtering?

A combination of intra-area dynamics and inter-area communication!

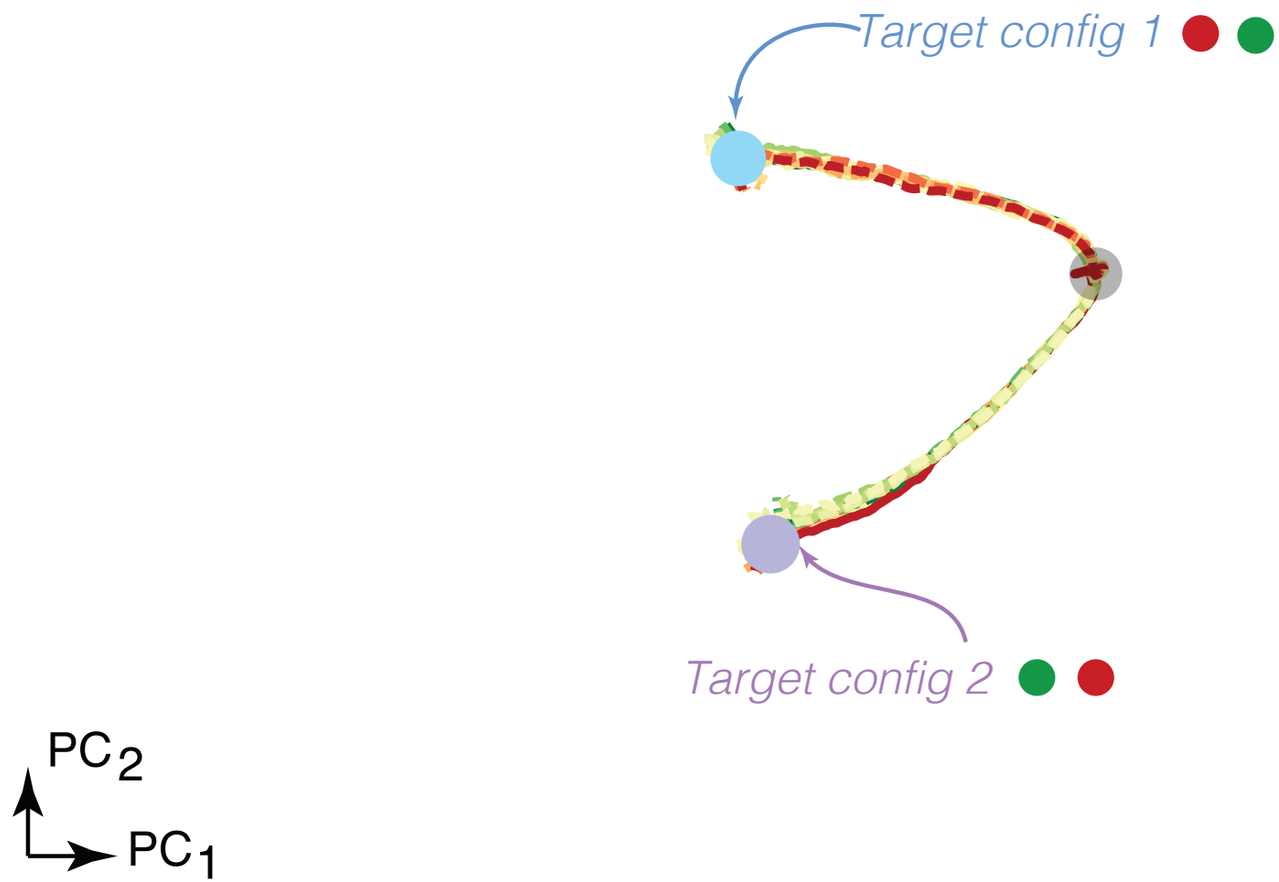
Dynamics in Area 1 orthogonalize direction and color information



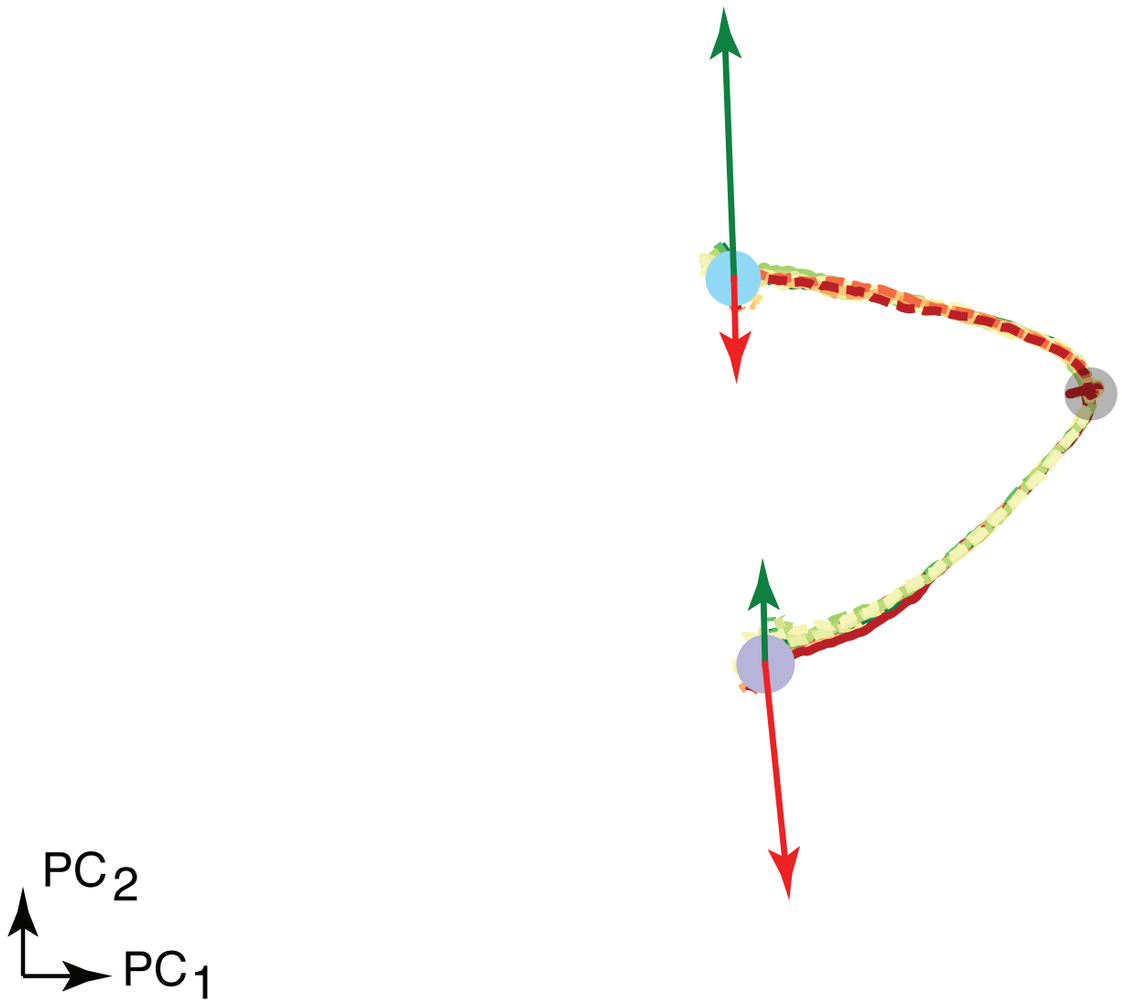
Dynamics in Area 1 orthogonalize direction and color information



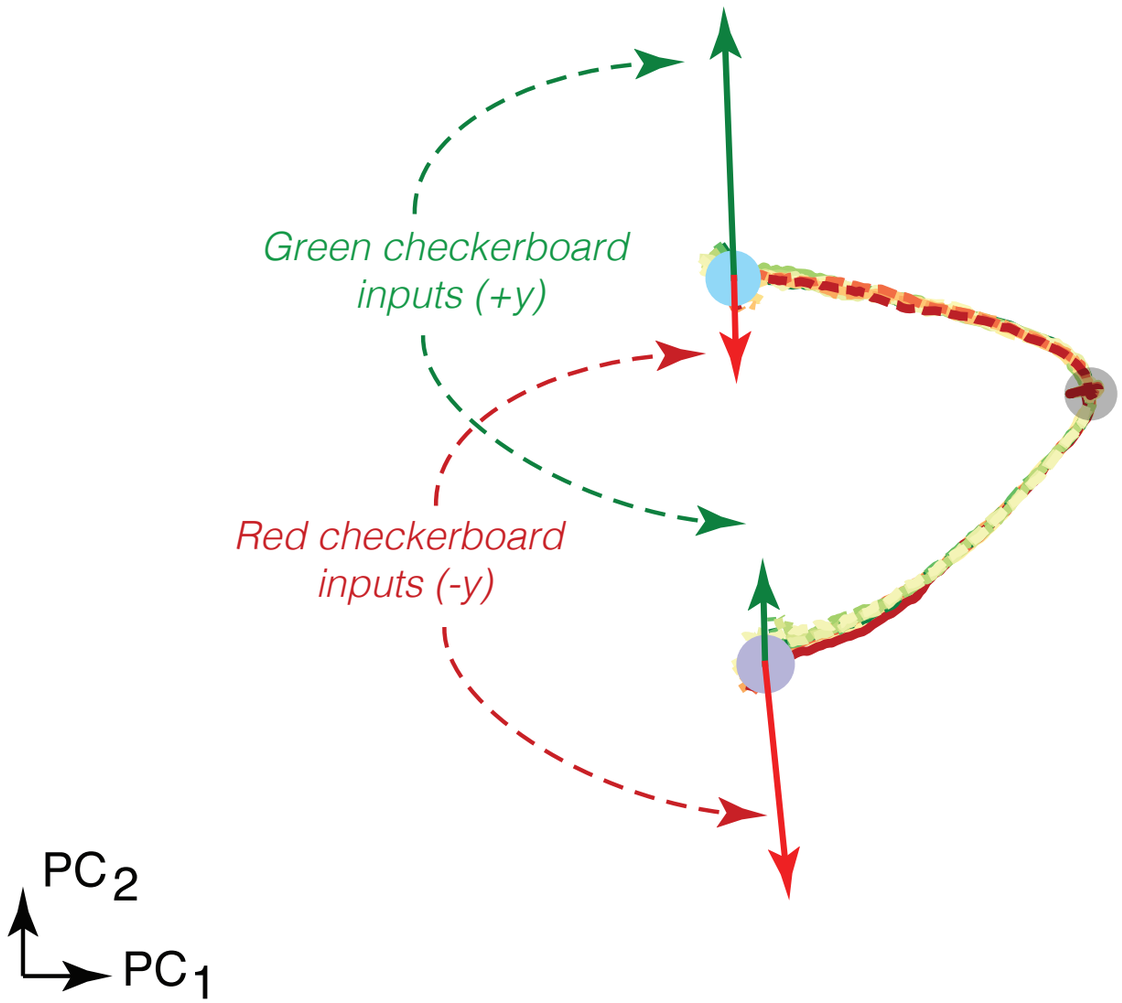
Dynamics in Area 1 orthogonalize direction and color information



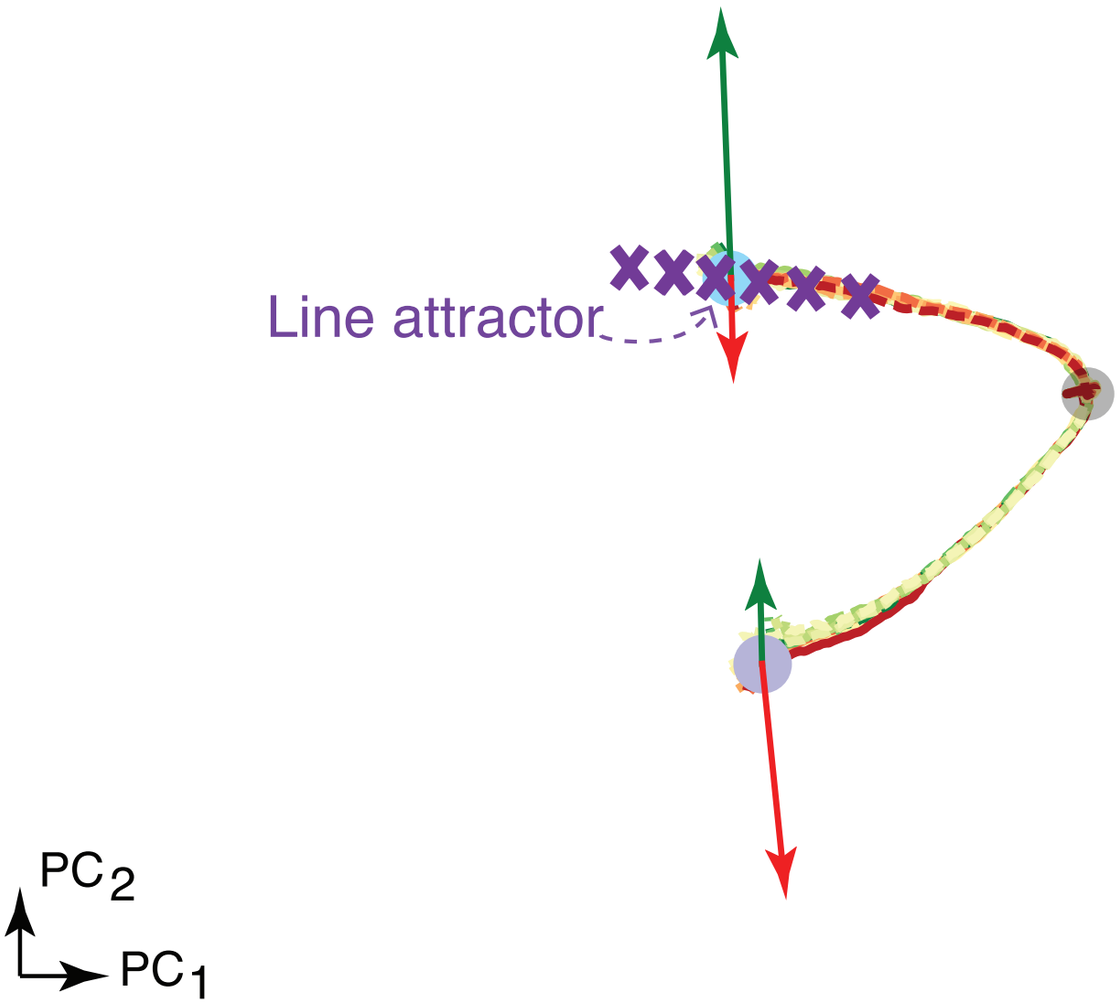
Dynamics in Area 1 orthogonalize direction and color information



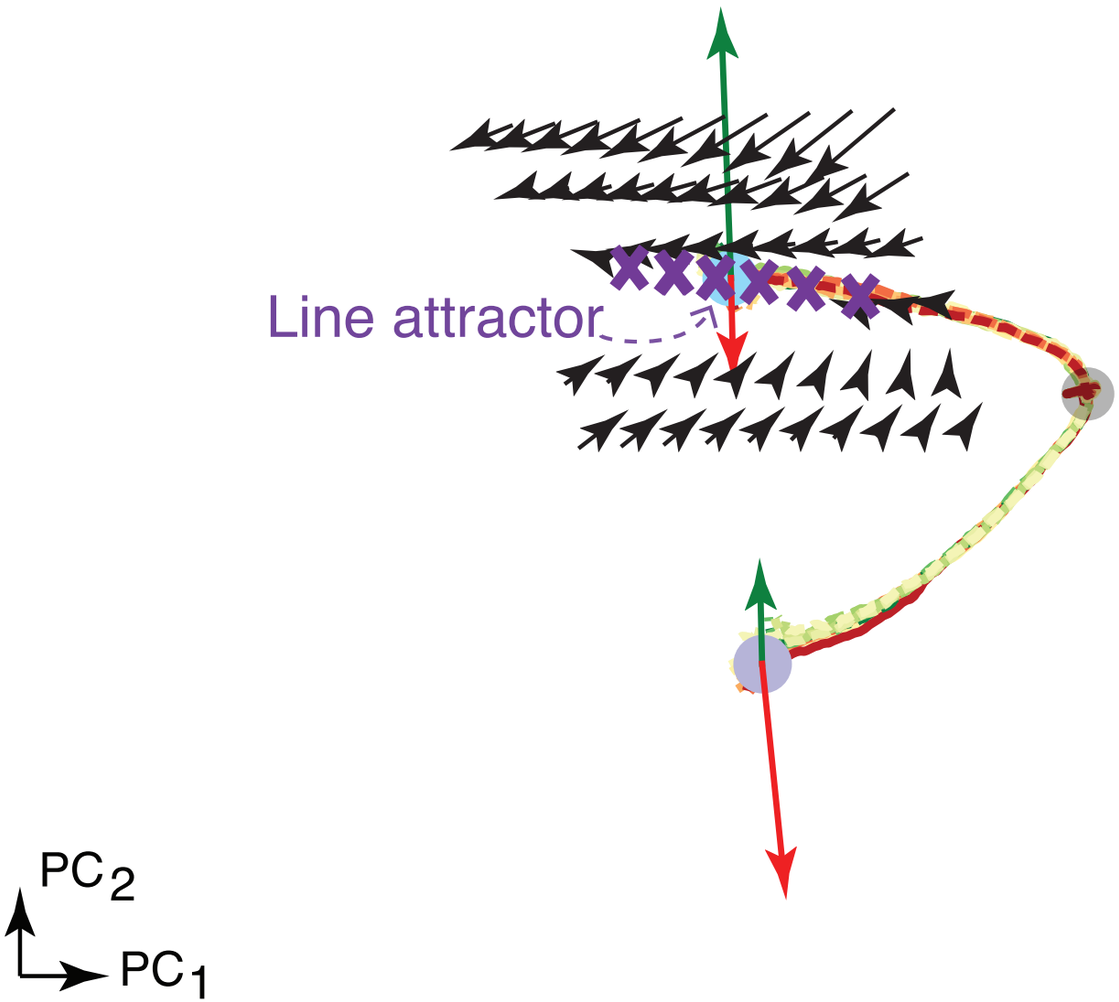
Dynamics in Area 1 orthogonalize direction and color information



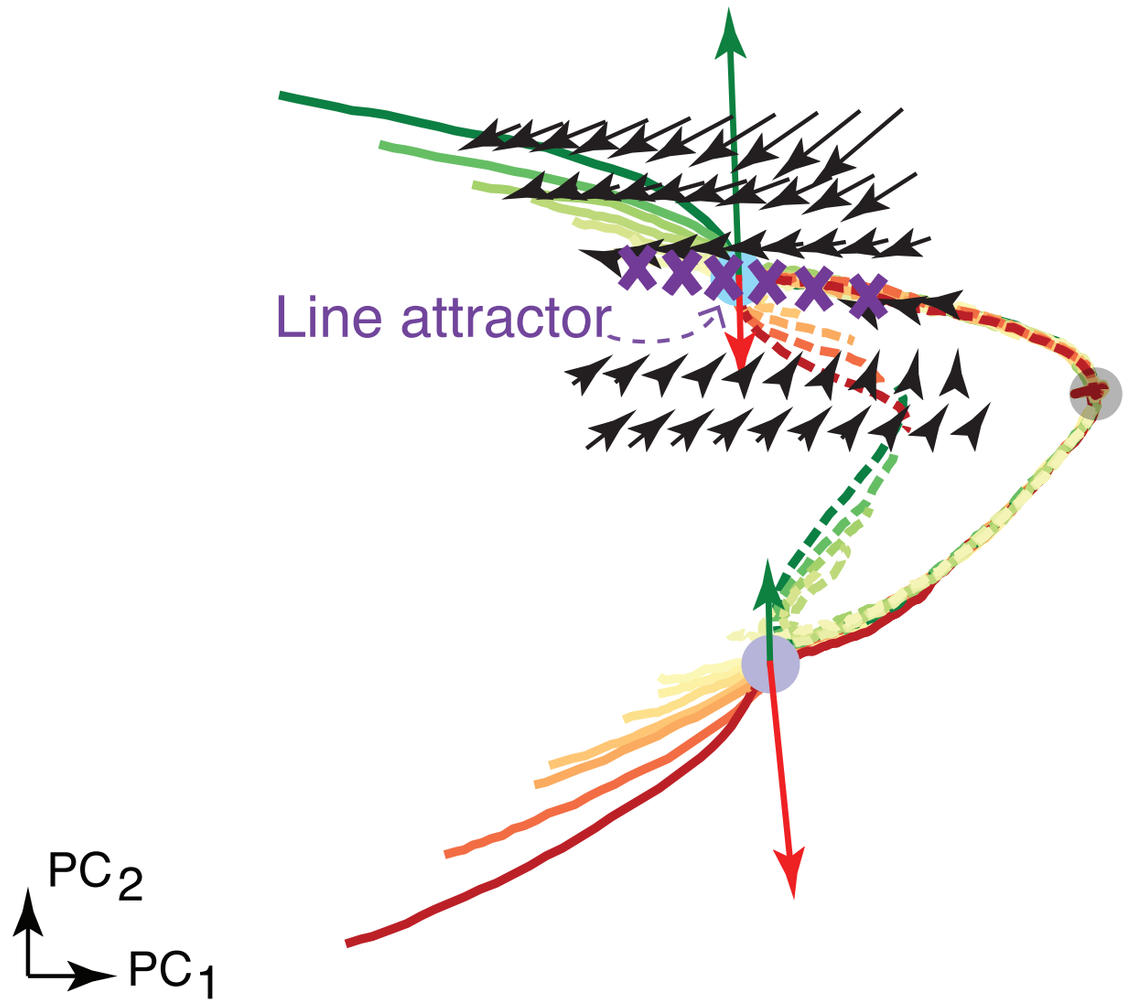
Dynamics in Area 1 orthogonalize direction and color information



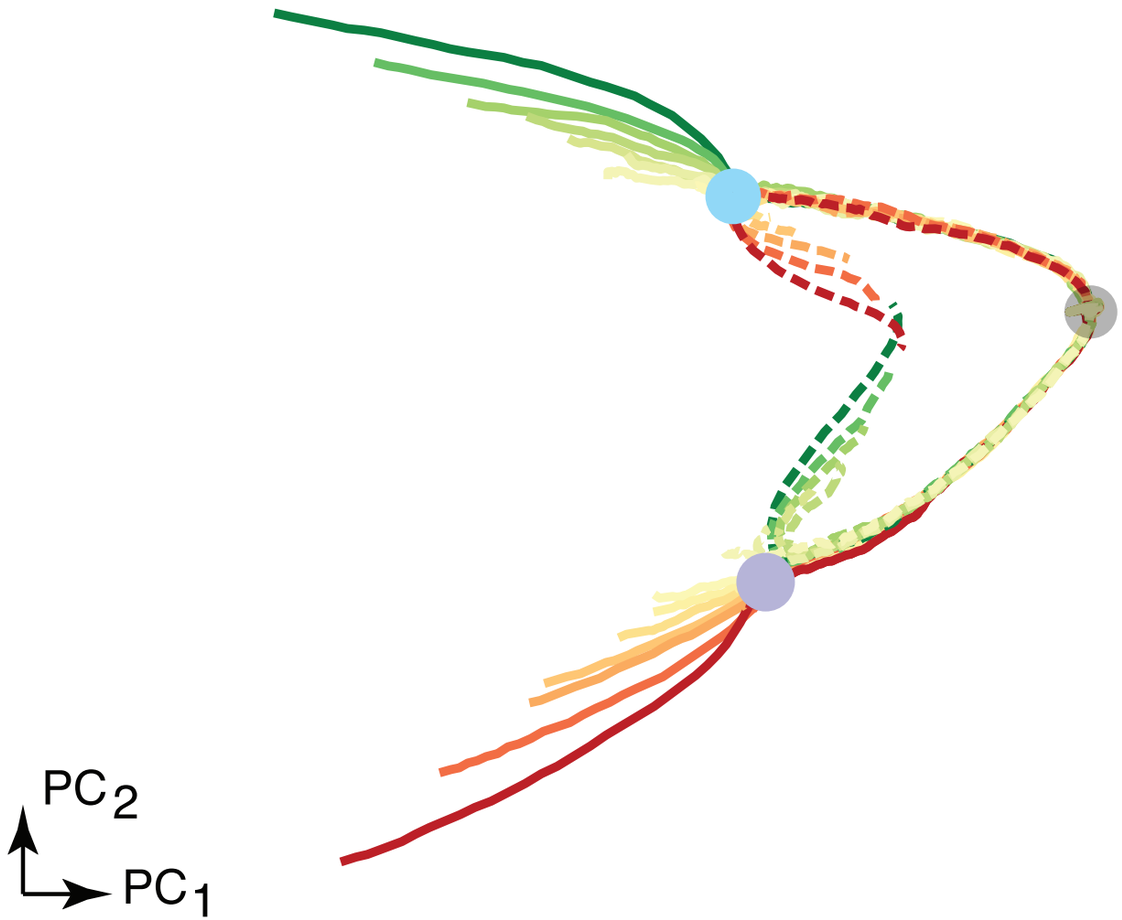
Dynamics in Area 1 orthogonalize direction and color information



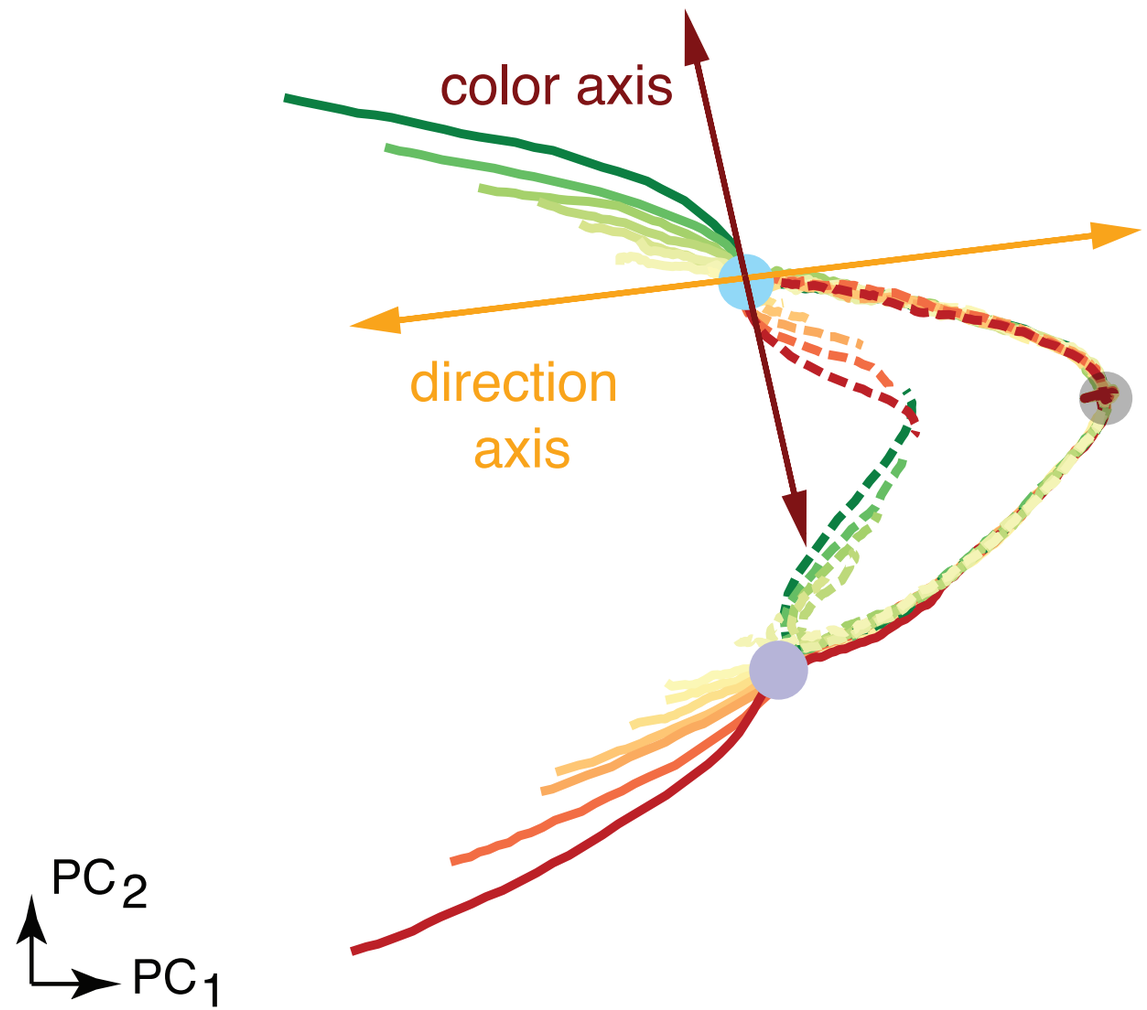
Dynamics in Area 1 orthogonalize direction and color information



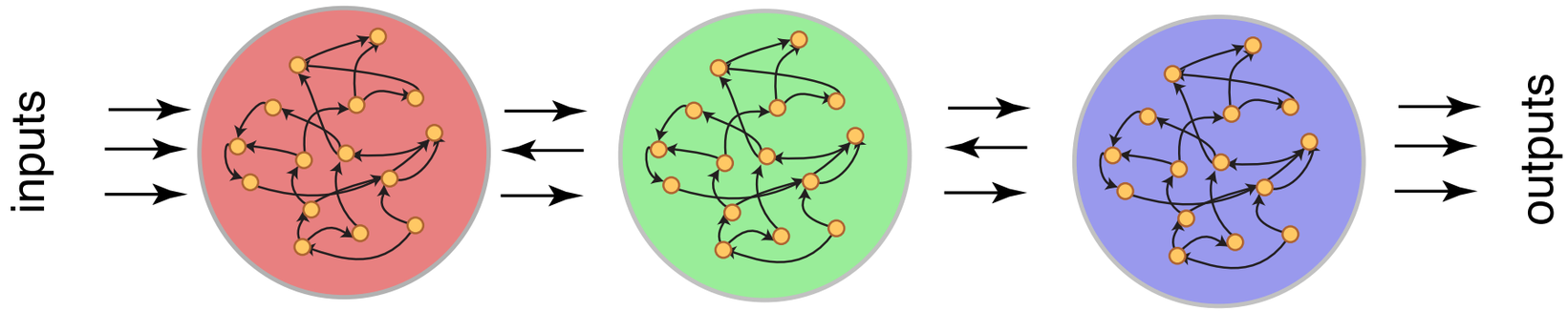
Dynamics in Area 1 orthogonalize direction and color information



Dynamics in Area 1 orthogonalize direction and color information



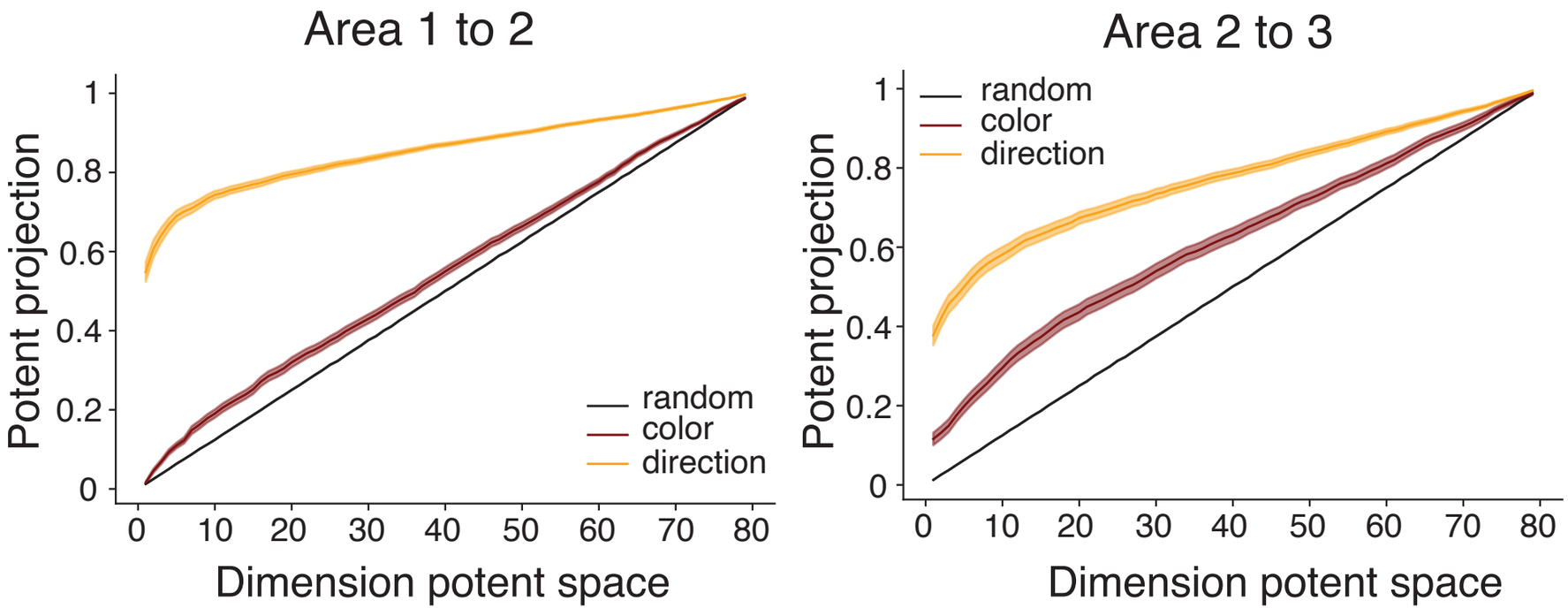
What about the computation to turn color into direction?



What mechanism achieves this color filtering?

A combination of intra-area dynamics and inter-area communication!

Direction information is then preferentially propagated



Some (non-exhaustive) ways ANNs and RNNs are used in other tasks

- **Navigation:** Banino, Barry et al., Nature 2018:

  Banino, Barry and colleagues trained agents (incorporating both CNNs and RNNs) that exhibited a grid-cell like code, and subsequently showed that this representation **improved navigation** in challenging environments.

- **Vision:** Yamins et al., PNAS 2014:

  Yamins and colleagues trained CNNs to perform image classification tasks, and observed deeper layers of the CNN explained more variance in deeper areas of the visual pipeline.

- **Multi-task**: Yang et al., Nature Neuroscience 2019:

  Yang and colleagues trained RNNs capable of doing 20 cognitive tasks, and characterized the compositionality and clustering of the artificial neurons.

- **Robustness to hyperparameters and architecture**: Maheswaranathan, Williams et al., NeurIPS 2019:

  Maheswaranathan, Williams, and colleagues swept hyperparameters and RNN architectures and suggested that while representations may differ based on hyperparameters, dynamical mechanisms are universal.

- **Working memory:** Orhan and Ma, Nature Neuroscience, 2019:

  Orhan and Ma show that persistent vs sequential solutions are task dependent.

- **Low-rank RNNs:** Mastrogiuseppe and Ostojic, Neuron 2018; Dubreuil et al., Nature Neuroscience 2022

- **Timing:** Remington et al., Neuron 2018.

Introduction take home points

- Artificial neural networks, though complex, are simpler than the brain.
- We can train RNNs to do the same tasks as behaving animals, and closely match both behavior and neural activity.
- RNNs can be probed to hypothesize mechanisms for how the brain does a computation.
- RNNs have wide applications in neuroscience.