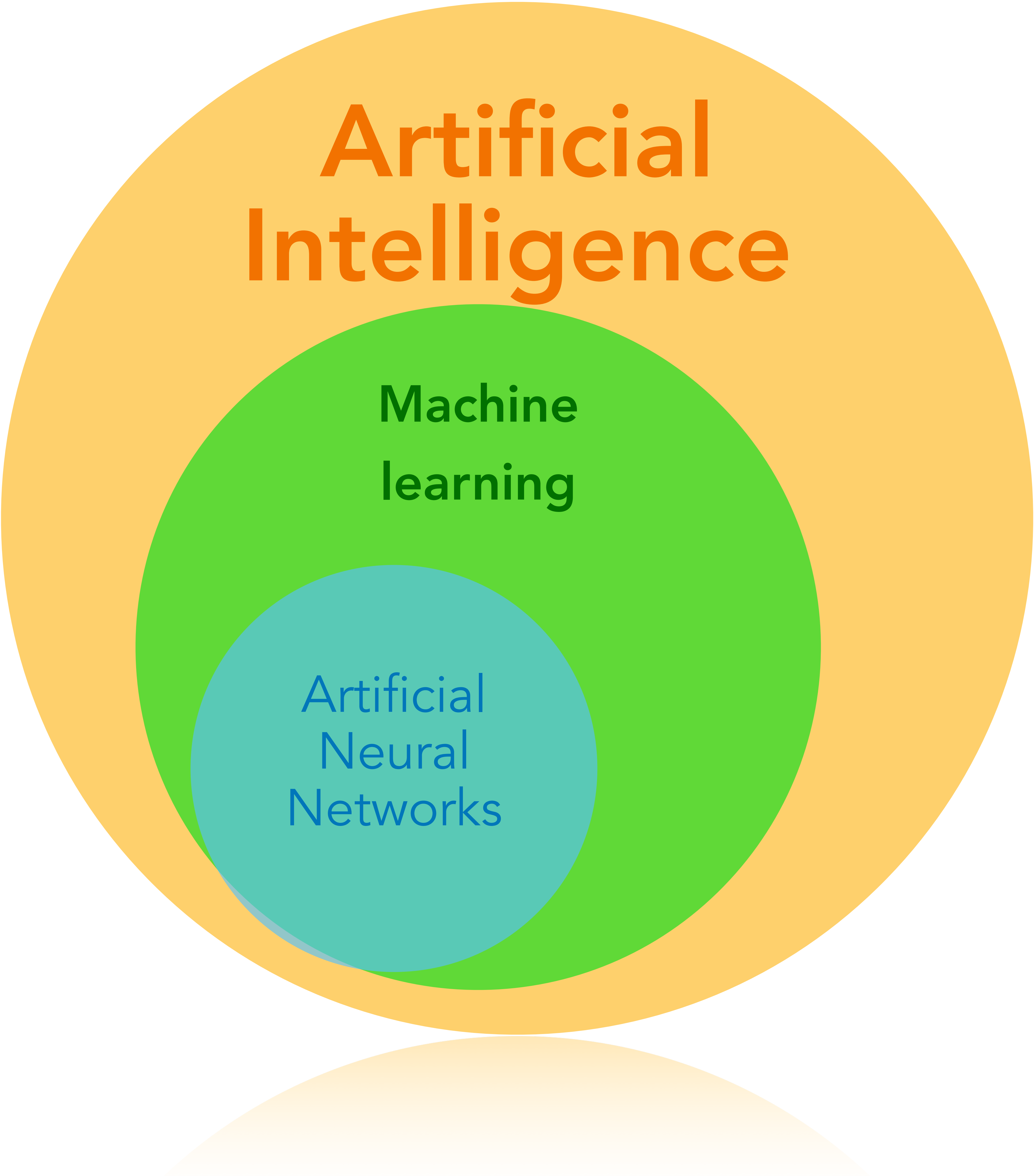
AI x Neuroscience

Course at ESPCI

Adrian Valente - adrian.valente@ens.psl.eu

AI x Neuroscience

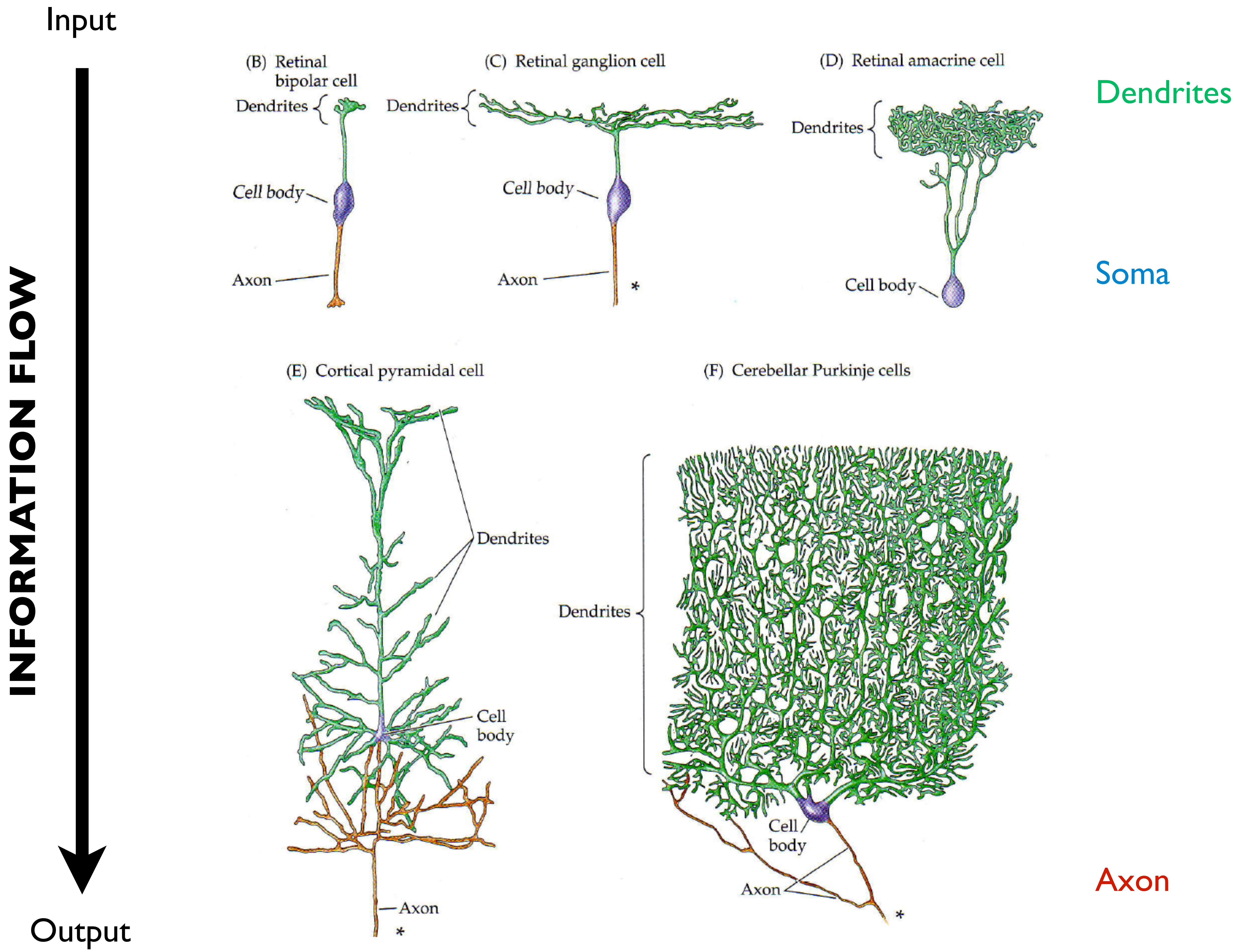


AI for neuroscience as a:

- **tool** (neuron tagging, spike sorting, statistical methods, DeepLabCut, …)
- **theory** (neural networks as computational models)

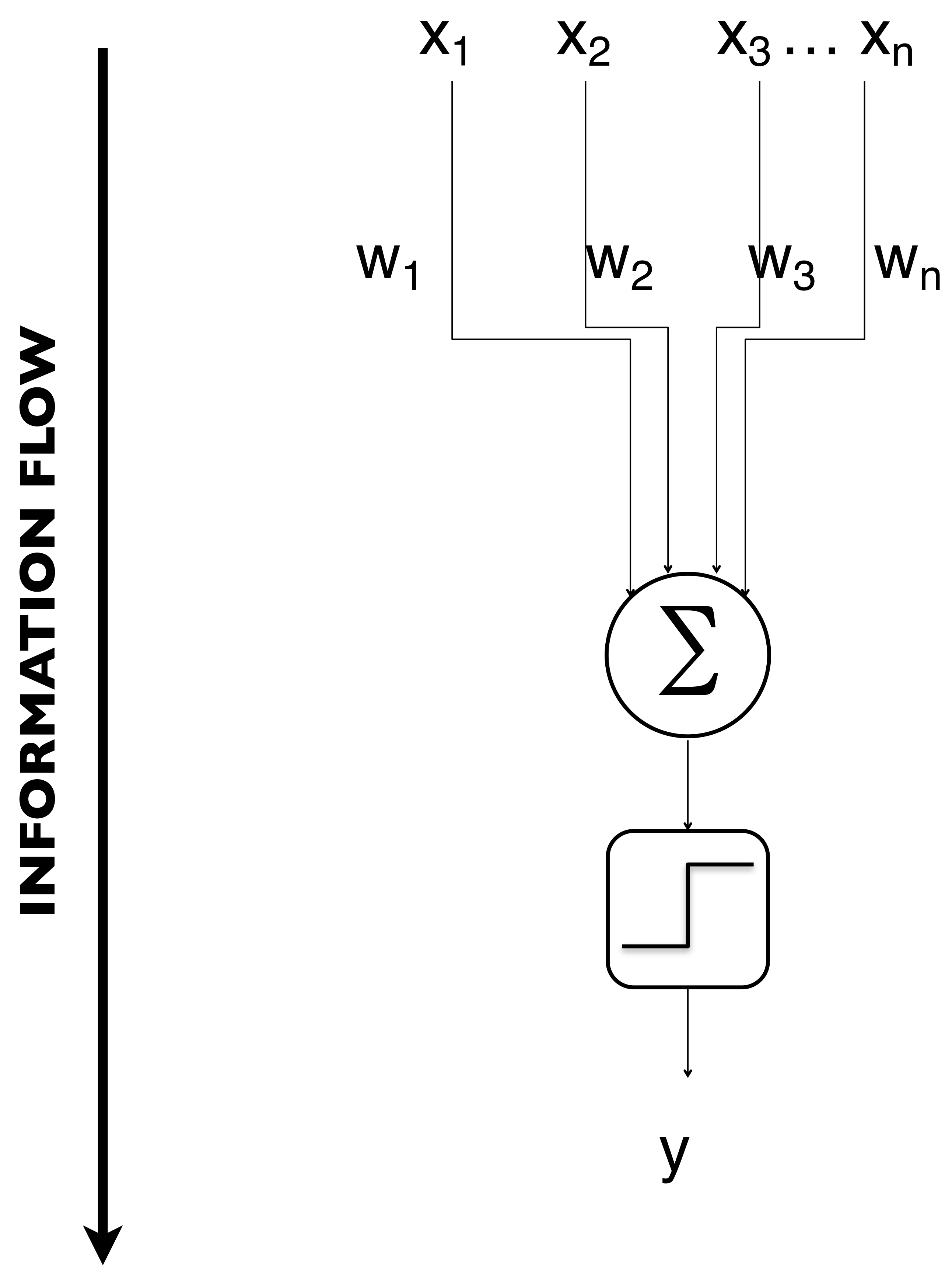
I - Brief history of ANNs

Real neurons



The “abstract" neuron

[McCulloch and Pitts (1943)]

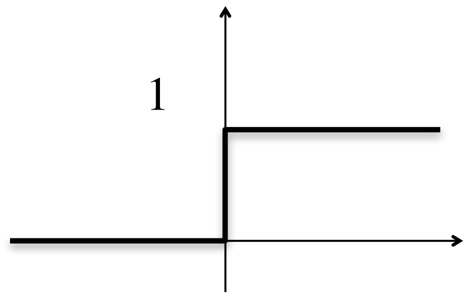


Synaptic Inputs

Synaptic weights

Summation $$\sum_{k=1}^{N} w_k x_k$$

Threshold



Output $$y = H\left(\sum_{k=1}^{N} w_k x_k - b\right)$$

The “abstract" neuron

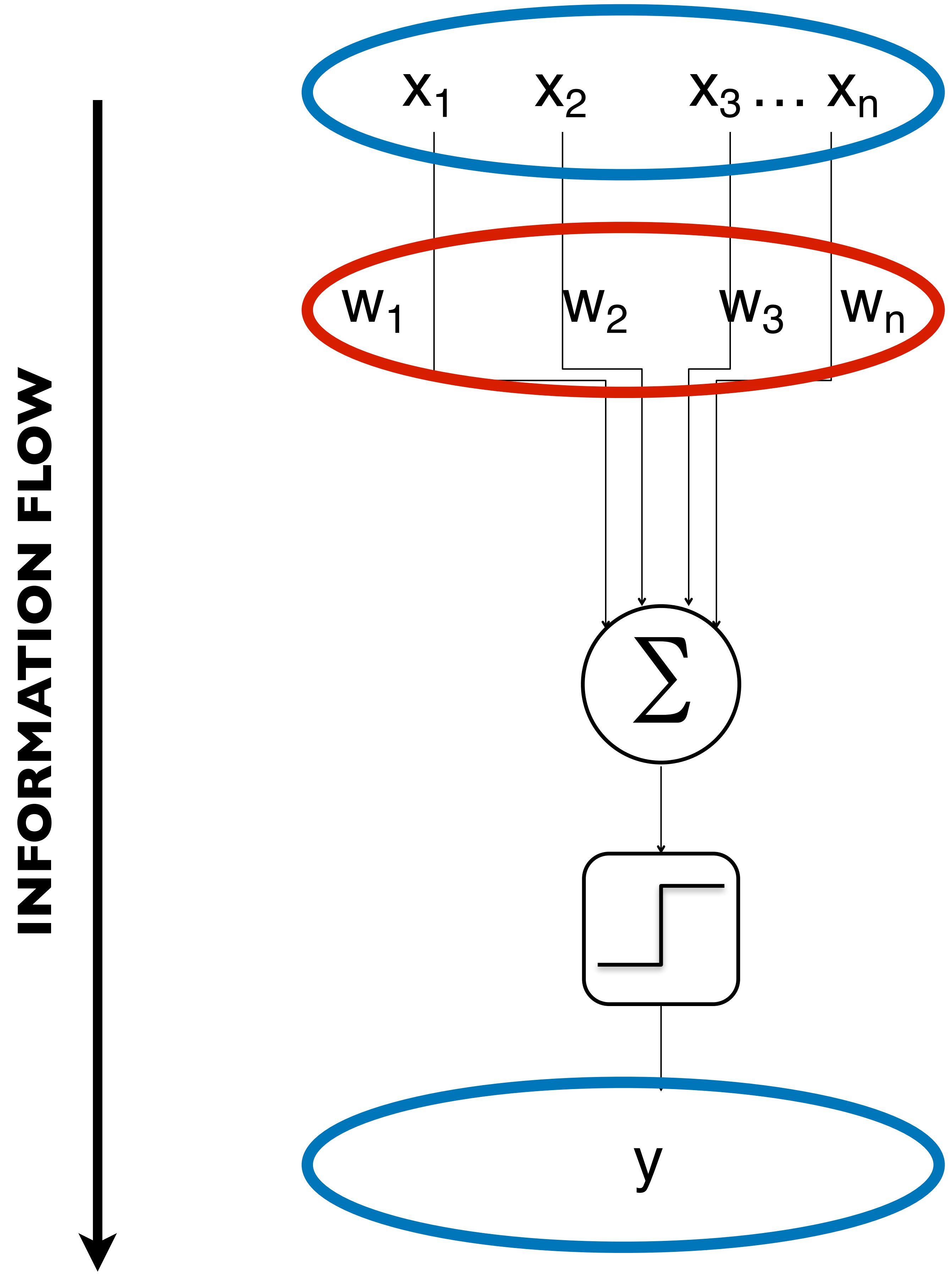
- No spikes
- No dendritic nonlinearities
- No neurotransmitter kinetics
- No genetic specificities (neuron classes)
- Synaptic delays can be added

Brain-like learning: the perceptron

[Rosenblatt (1958)]

Frank Rosenblatt

given — adjust — given

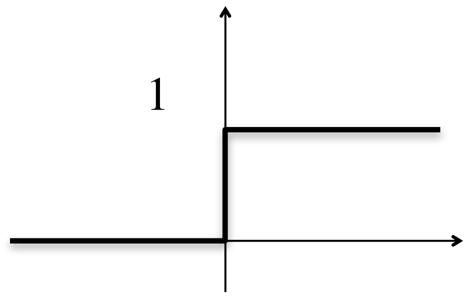


Synaptic Inputs

Synaptic weights

Summation $\sum_{k=1}^{N} w_k x_k$

Threshold

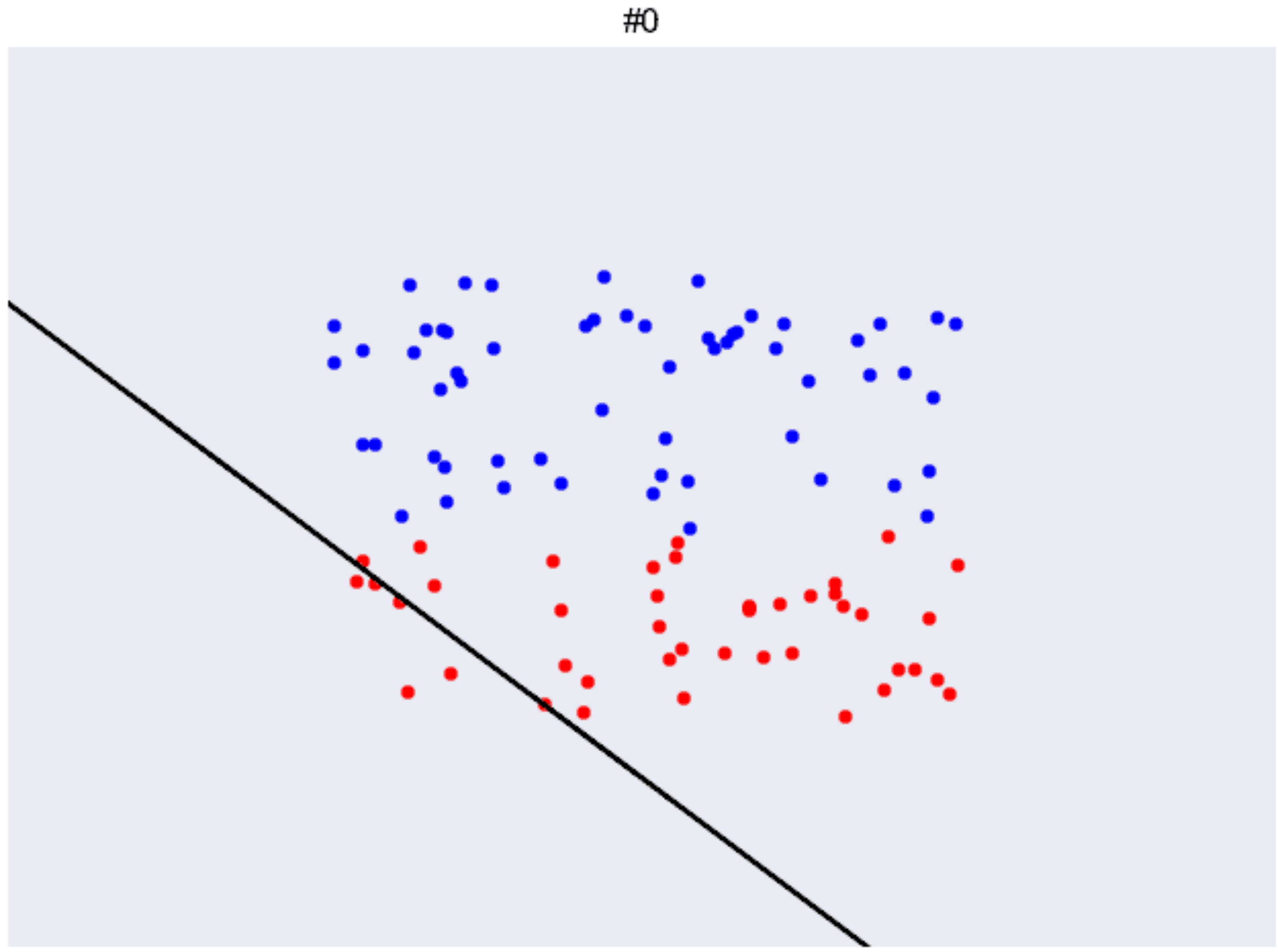


Output $y = H\left(\sum_{k=1}^{N} w_k x_k - b\right)$

Brain-like learning: the perceptron

[Rosenblatt (1958)]

$$y = H\left(\sum_{k=1}^{N} w_k x_k - b\right)$$ (where y=0 or 1)



More specifically:

$$y = 1 \quad \text{if} \quad \sum w_k x_k - b > 0$$

$$y = \quad \text{if} \quad \sum w_k x_k - b < 0$$

Learning: iterative procedure to update $\{w_k\}$

Brain-like learning: the perceptron

[New York Times (1958)]

Electronic 'Brain' Teaches Itself

The Navy last week demonstrated the embryo of an electronic computer named the Perceptron which, when completed in about a year, is expected to be the first non-living mechanism able to "perceive, recognize and identify its surroundings without human training or control." Navy officers demonstrating a preliminary form of the device in Washington said they hesitated to call it a machine because it is so much like a "human being without life."

Dr. Frank Rosenblatt, research psychologist at the Cornell Aeronautical Laboratory, Inc., Buffalo, N. Y., designer of the Perceptron, conducted the demonstration. The machine, he said, would be the first electronic device to think as the human brain. Like humans, Perceptron will make mistakes at first, "but it will grow wiser as it gains experience," he said.

The first Perceptron, to cost about $100,000, will have about 1,000 electronic "association cells" receiving electrical impulses from an eyelike scanning device with 400 photocells. The human brain has ten billion responsive cells, including 100,000,000 connections with the eye.

Difference Recognized

recognize the difference between right and left, almost the way a child learns.

When fully developed, the Perceptron will be designed to remember images and information it has perceived itself, whereas ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons, Dr. Rosenblatt said, will be able to recognize people and call out their names. Printed pages, longhand letters and even speech commands are within its reach. Only one more step of development, a difficult step, he said, is needed for the device to hear speech in one language and instantly translate it to speech or writing in another language.

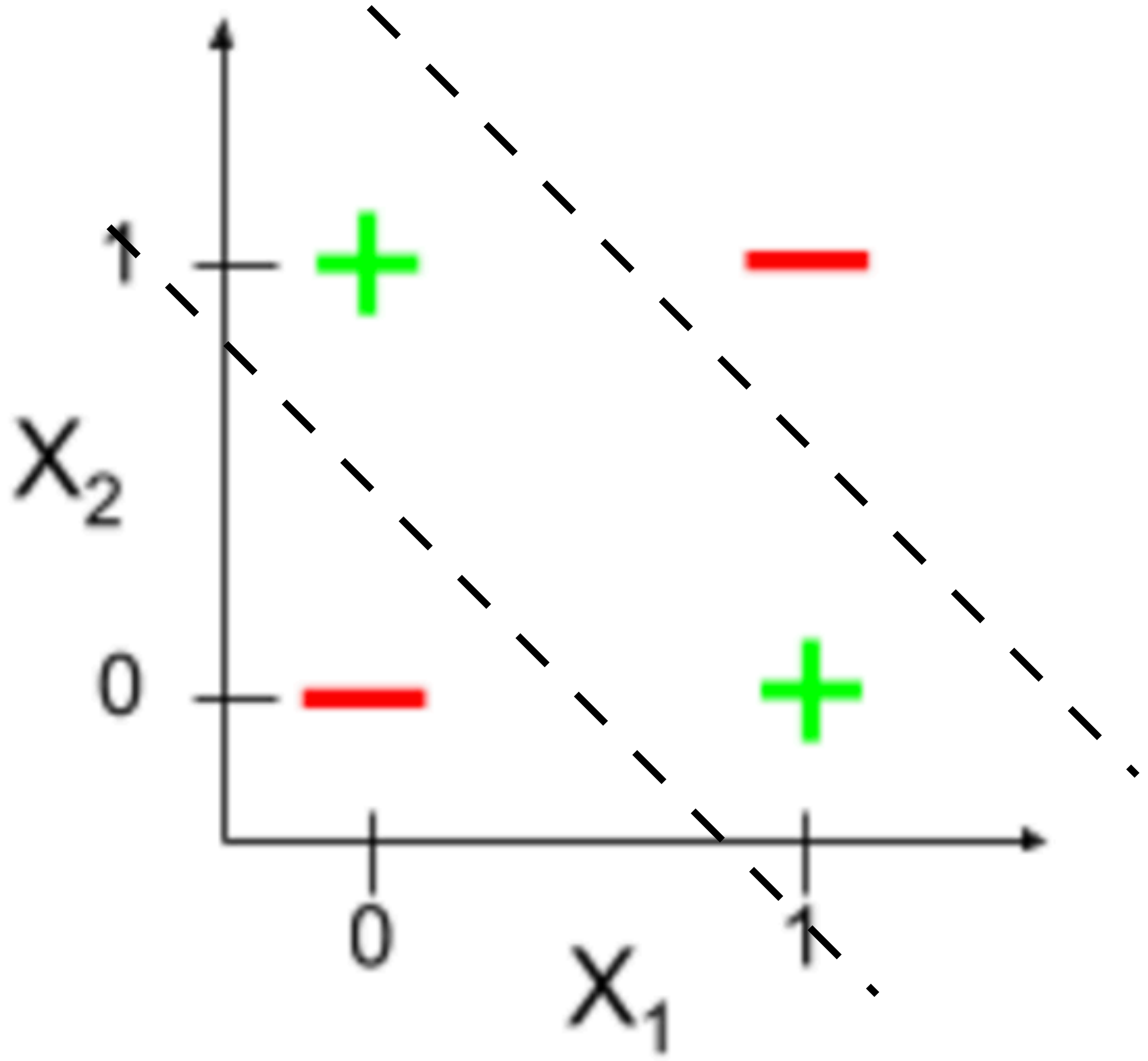
Self-Reproduction

In principle, Dr. Rosenblatt said, it would be possible to build Perceptrons that could reproduce themselves on an assembly line and which would be "conscious" of their existence.

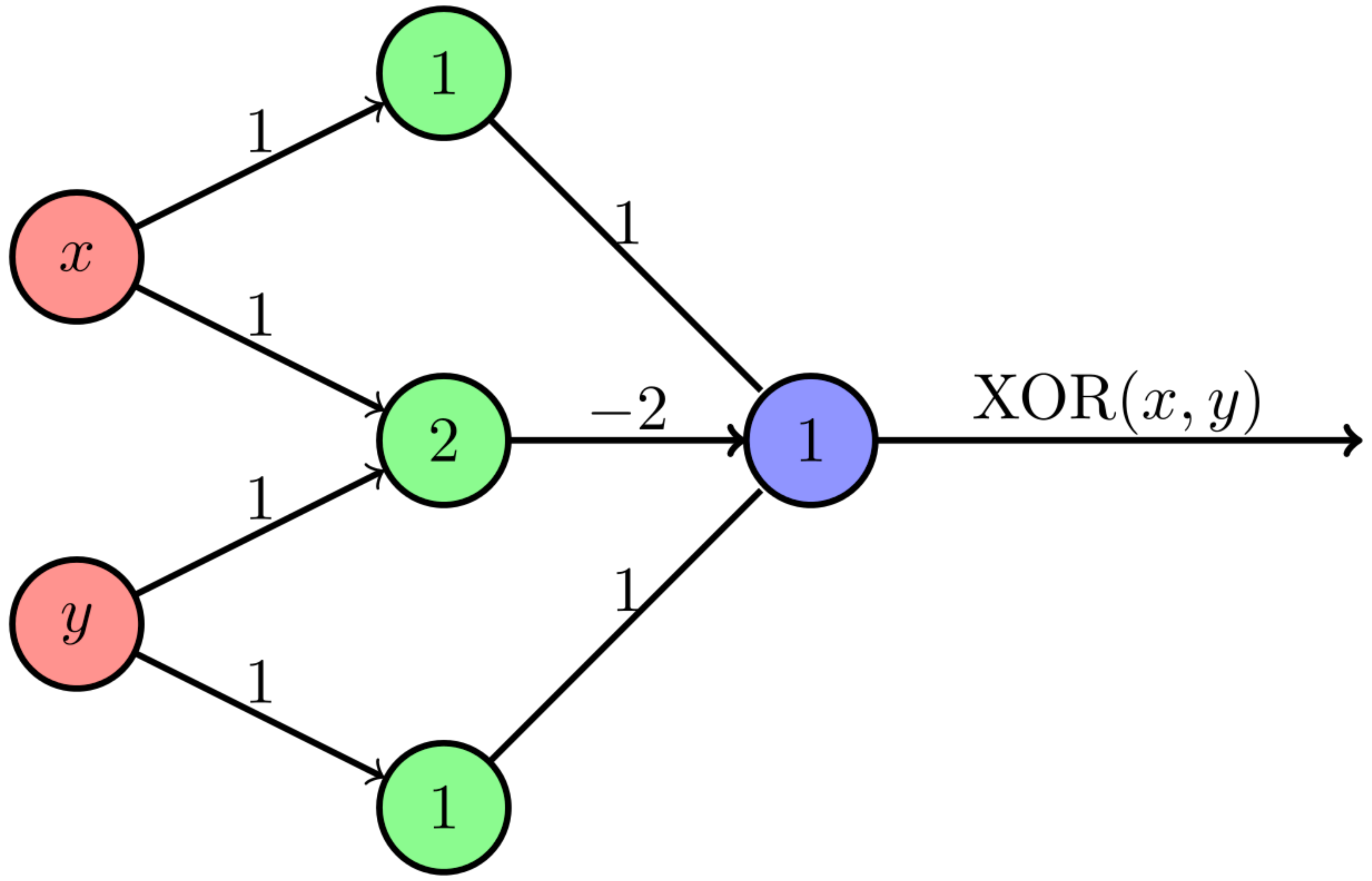
Perceptron, it was pointed out, needs no "priming." It is not necessary to introduce it to surroundings and circumstances, record the data involved and then store them for future comparison as is the case

The XOR problem

[Minsky & Papert, *Perceptrons*, (1969)



Possible solution: multi-layer perceptron



but how to train them ?

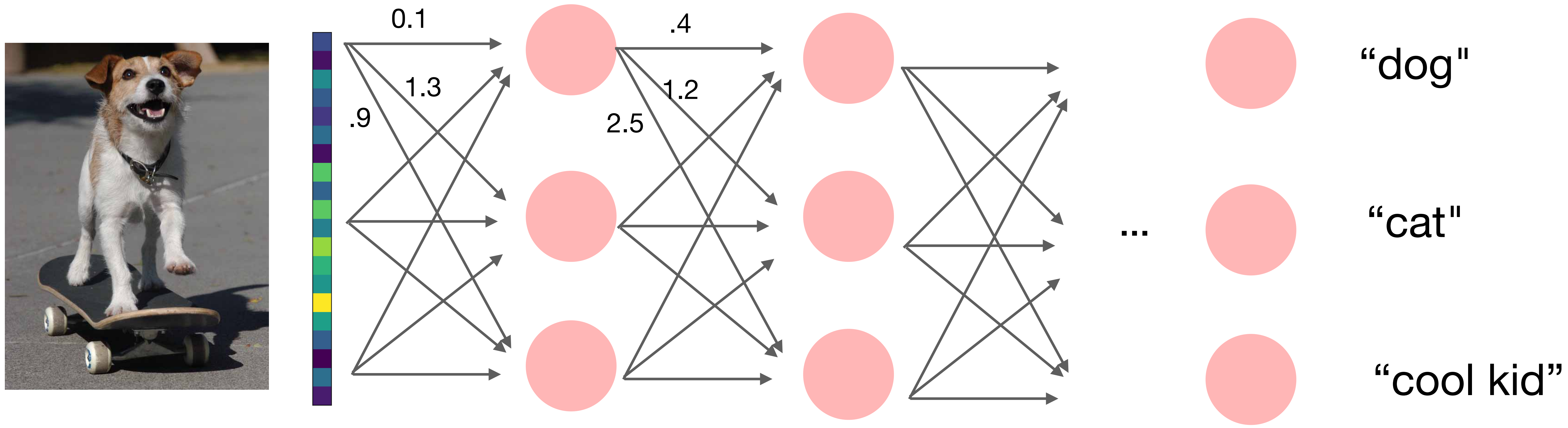
AI winter…

(more of an ANN winter)

Backpropagation algorithm (feedforward nets)

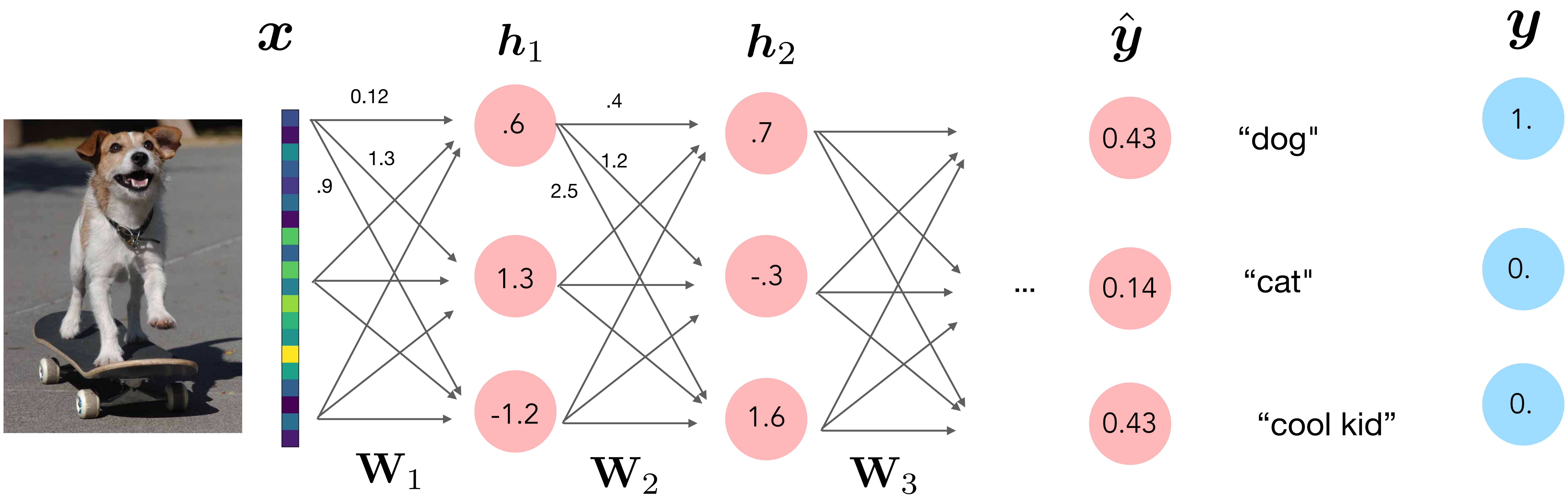
Learning representations by back-propagating errors (1986)

David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*



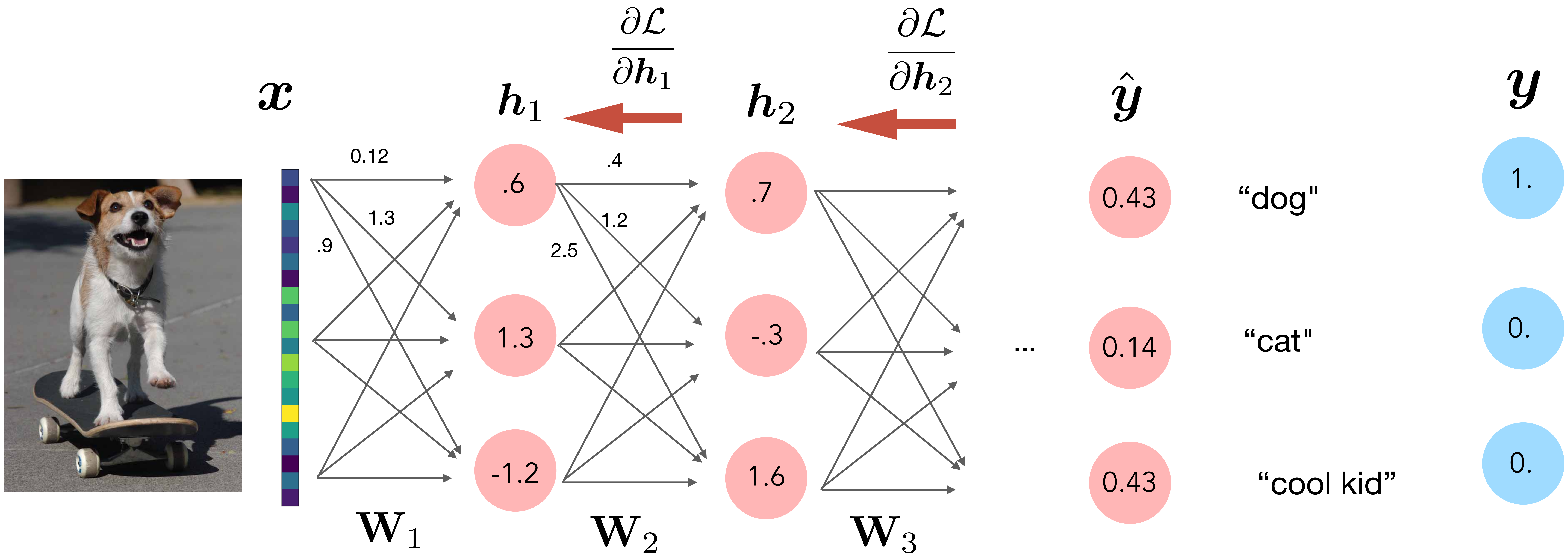
Goal: learn mapping vector -> vector

Backpropagation algorithm (feedforward nets)



Error or "loss": $\mathcal{L} = (y - \hat{y})^2$

Backpropagation algorithm (feedforward nets)

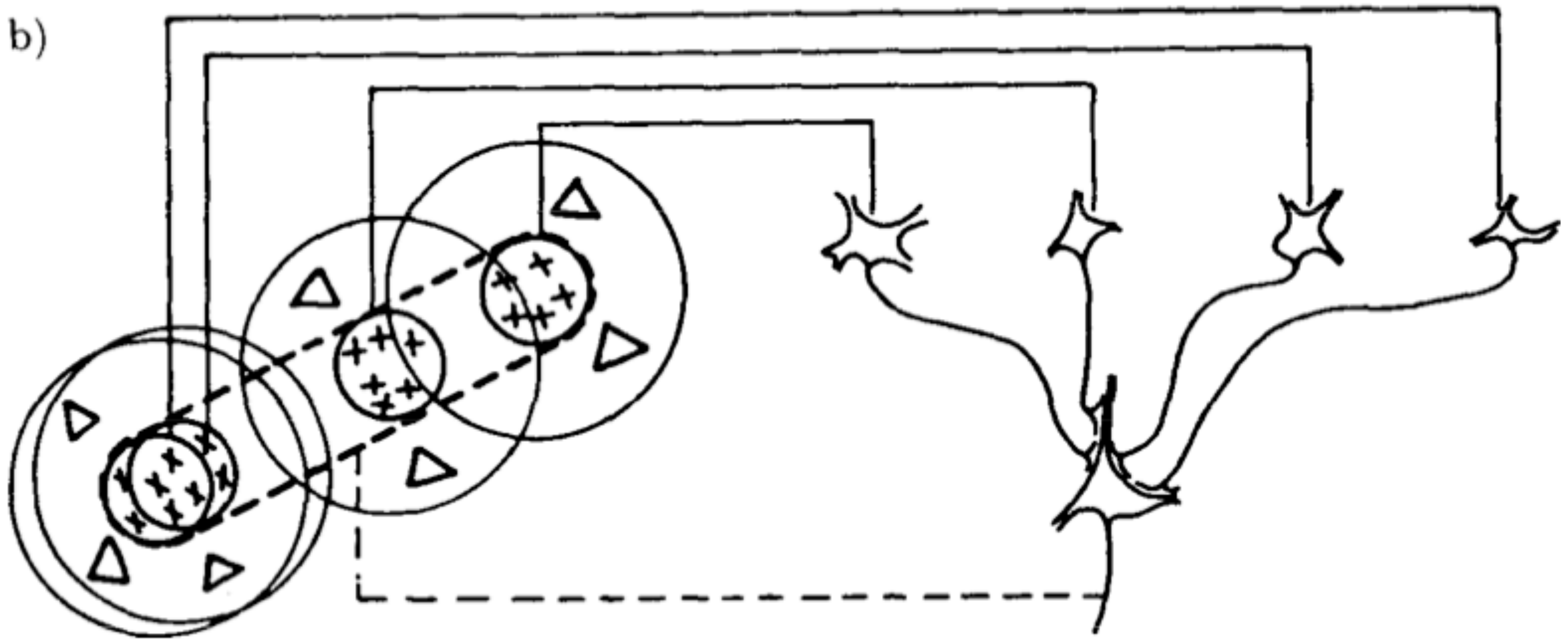


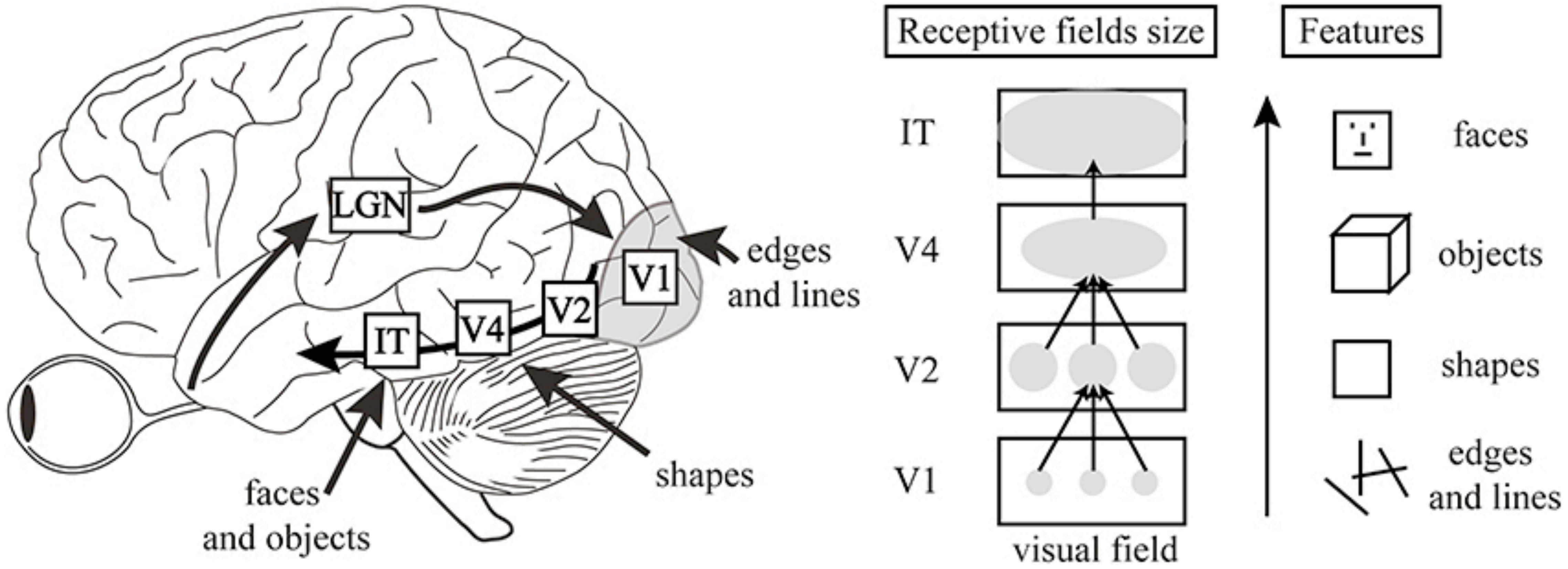
Error or “loss”: $\mathcal{L} = (y - \hat{y})^2$

Gradient descent update: $\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{w}}$

Convolutional neural networks

David Hubel & Torsten Wiesel (Nobel prize 1981)





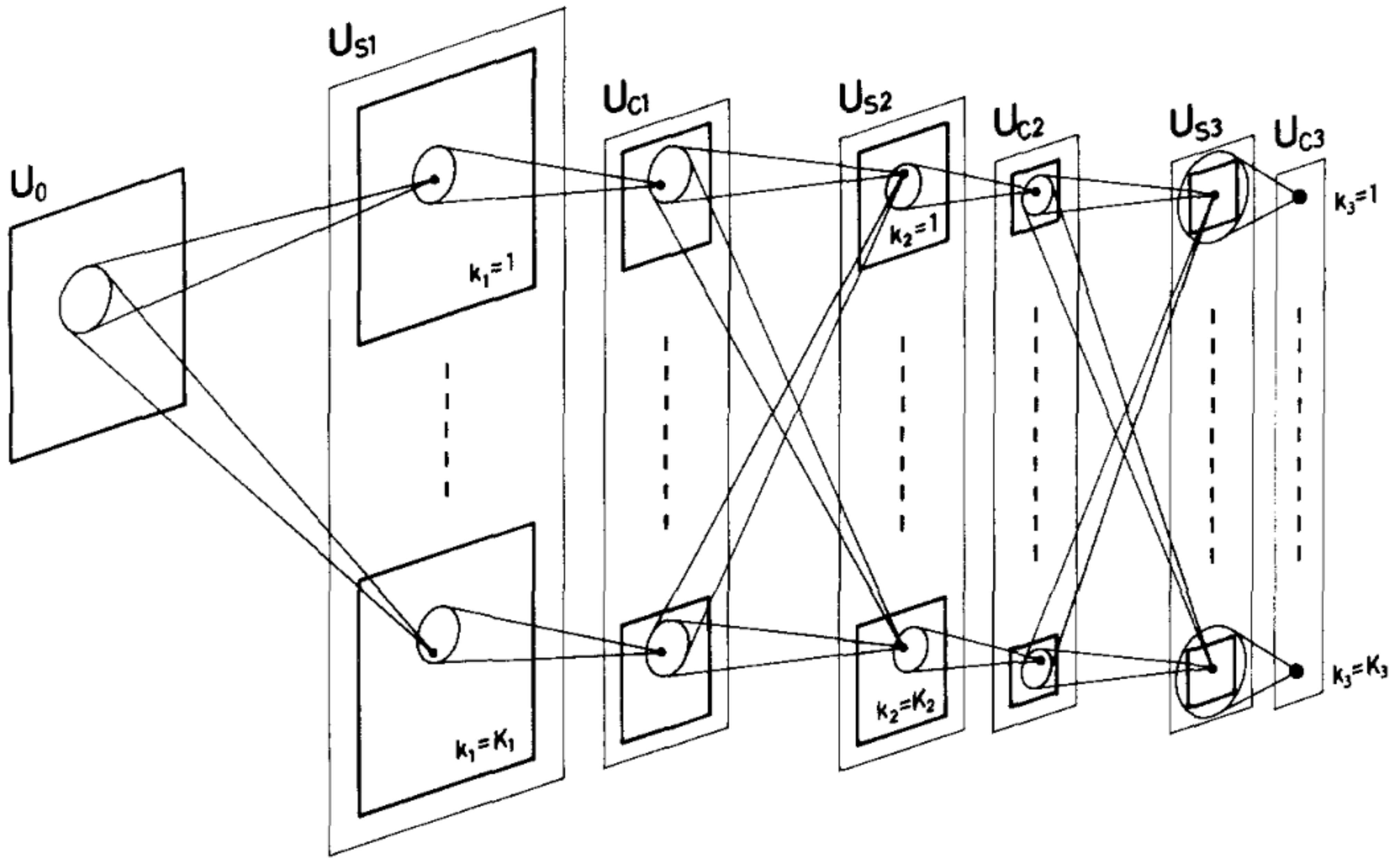
Convolutional neural networks

Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Kunihiko Fukushima

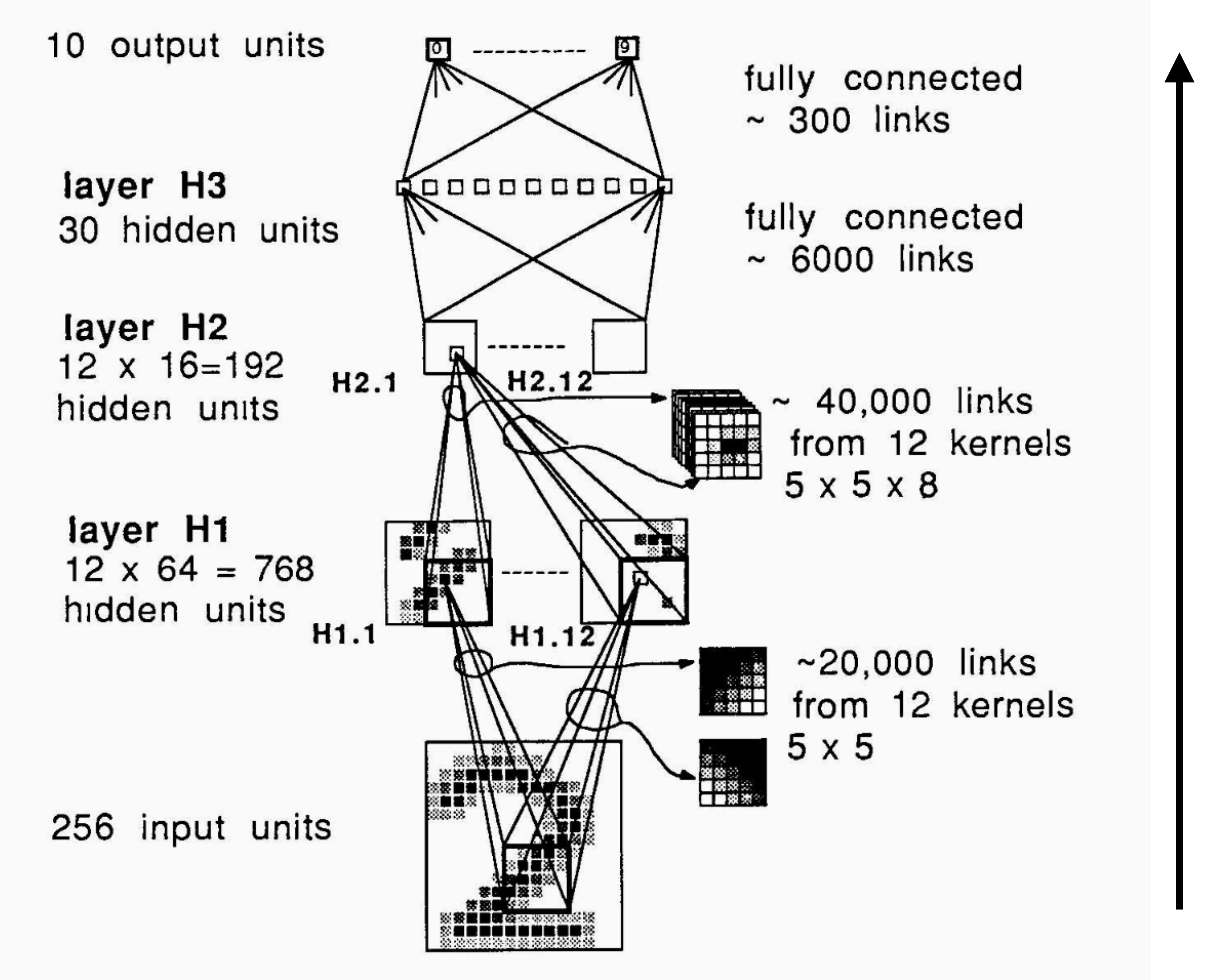
NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan

(1980)



Backpropagation Applied to Handwritten Zip Code Recognition

LeCun et al., (1989)



… and the ANN explosion!

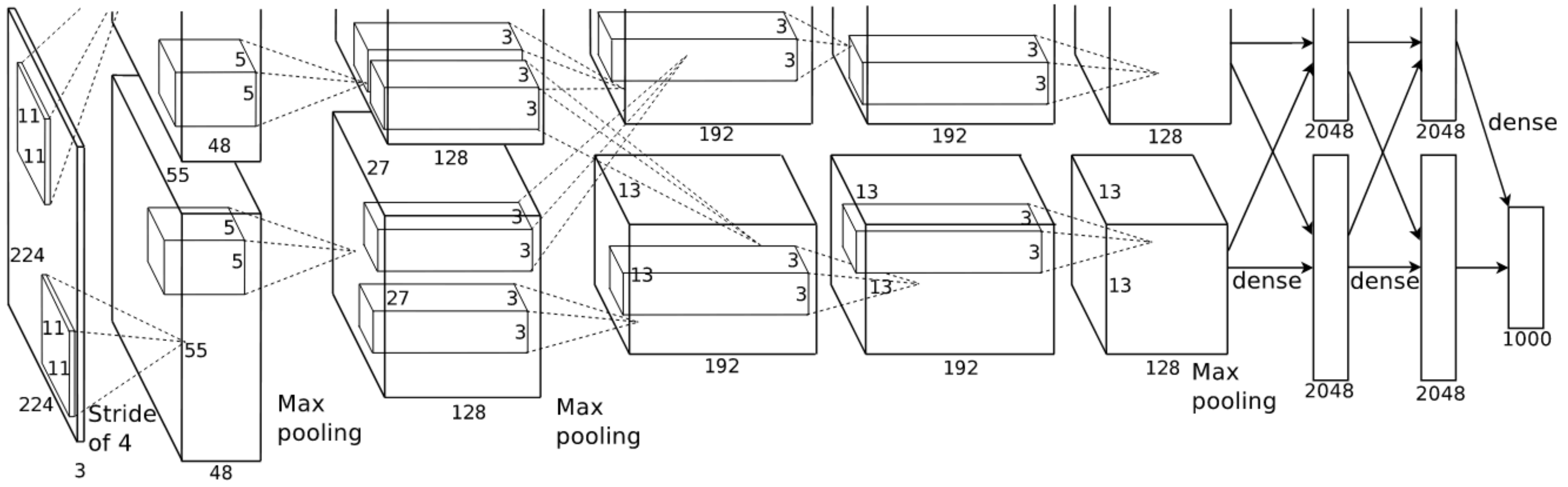
ImageNet Classification with Deep Convolutional Neural Networks

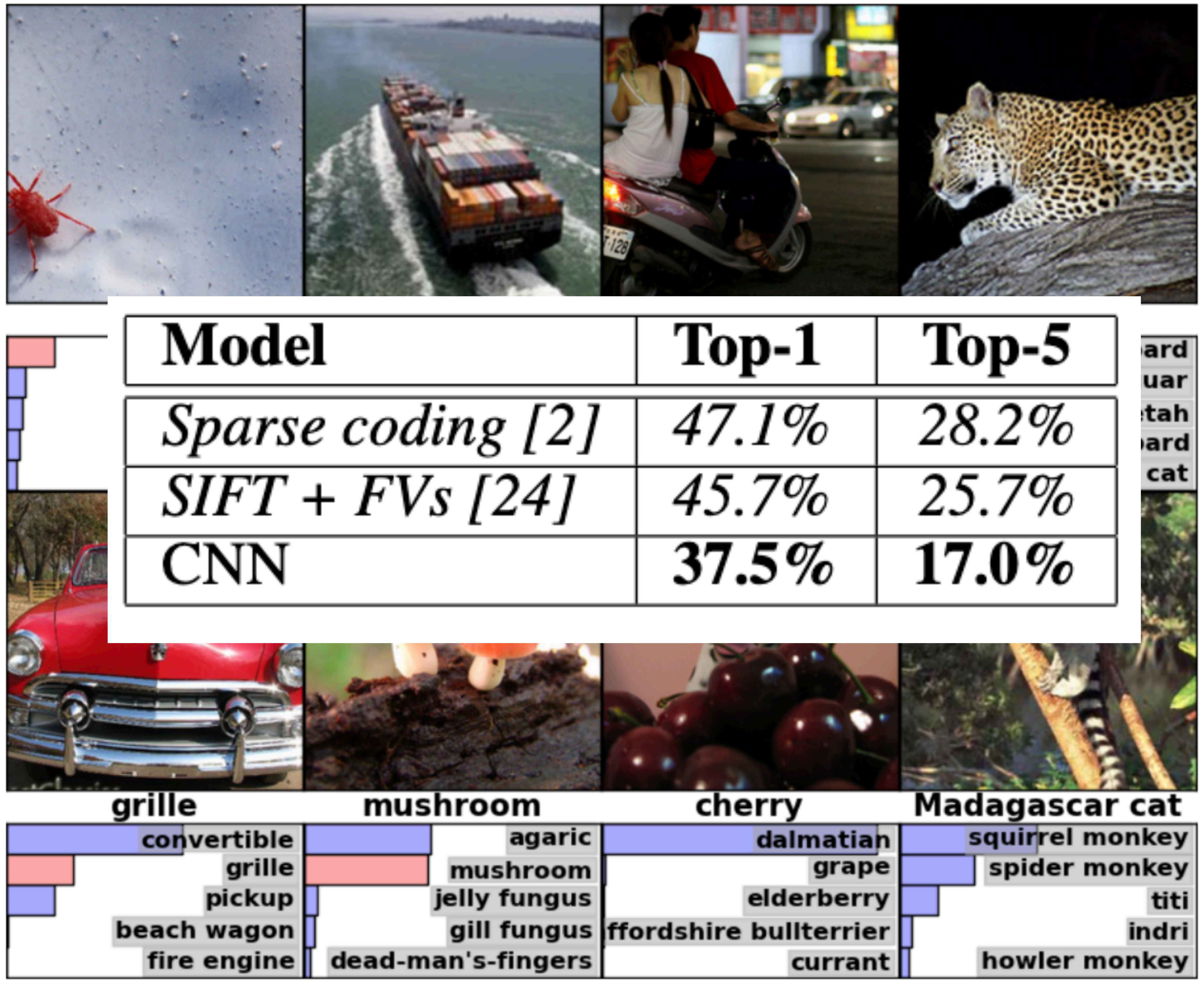
Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

(2012)





Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

Meanwhile in theory labs

Approximation by Superpositions of a Sigmoidal Function*

G. Cybenko†

Abstract. In this paper we demonstrate that finite linear combinations of compositions of a fixed, univariate function and a set of affine functionals can uniformly approximate any continuous function of n real variables with support in the unit hypercube; only mild conditions are imposed on the univariate function. Our results settle an open question about representability in the class of single hidden layer neural networks. In particular, we show that arbitrary decision regions can be arbitrarily well approximated by continuous feedforward neural networks with only a single internal, hidden layer and any continuous sigmoidal nonlinearity. The paper discusses approximation properties of other possible types of nonlinearities that might be implemented by artificial neural networks.

Key words. Neural networks, Approximation, Completeness.

(1989)

Meanwhile in theory labs

Universal approximation theorem

A feedforward neural network can, with sufficient hidden neurons and the correct set of weights, approximate any function $f : \mathbb{R}^n \to \mathbb{R}^m$ (of compact support)

Remaining problem: variable-length inputs

So far we consider inputs that can be mapped to vectors in $\mathbb{R}^n$

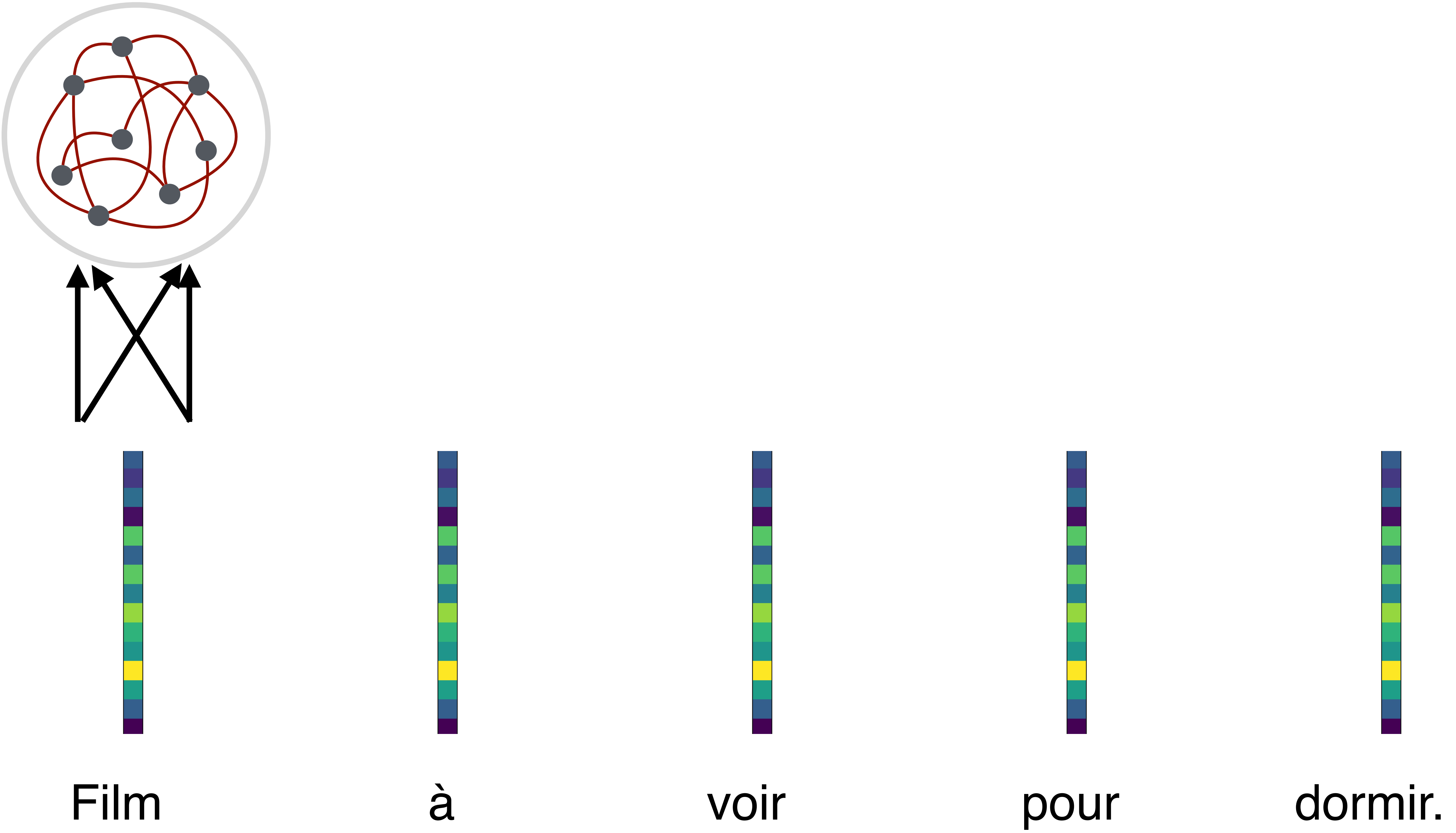
What about:

- text
- sound
- video
- time series…

Recurrent neural networks (RNNs)

Objective: learn a mapping sequence -> vector

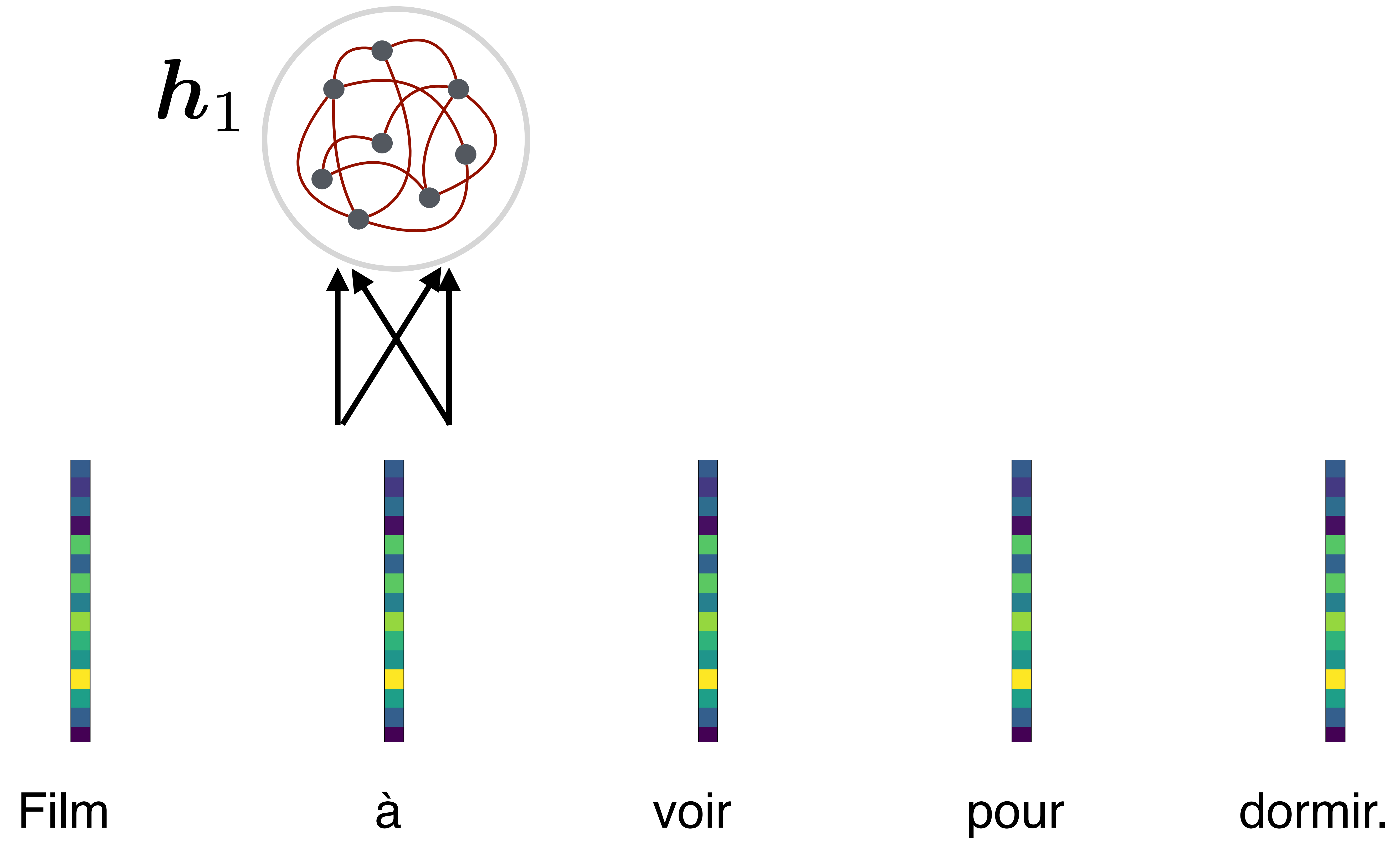
Example: sentiment analysis



Recurrent neural networks (RNNs)

Objective: learn a mapping sequence -> vector

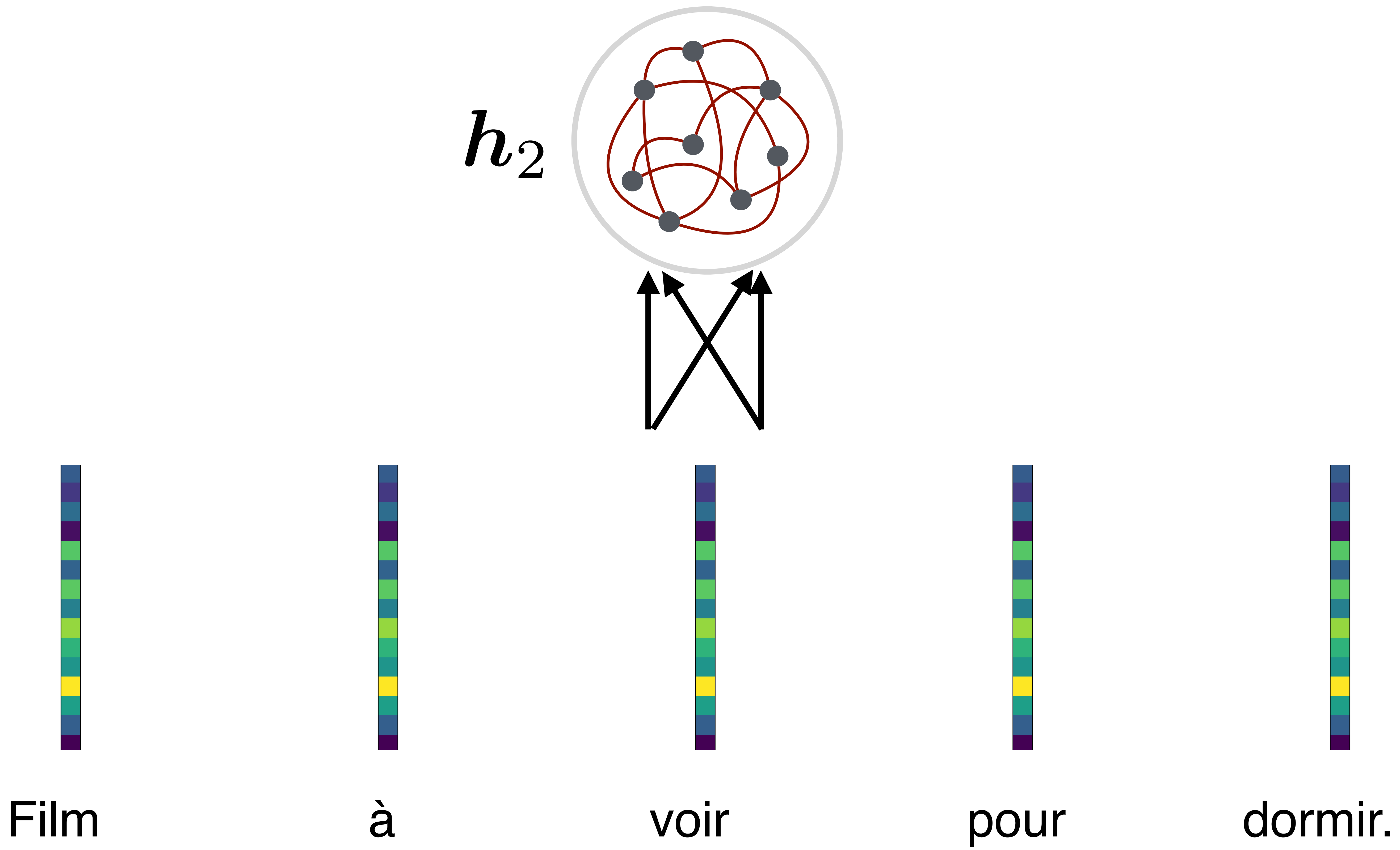
Example: sentiment analysis



Recurrent neural networks (RNNs)

Objective: learn a mapping sequence -> vector

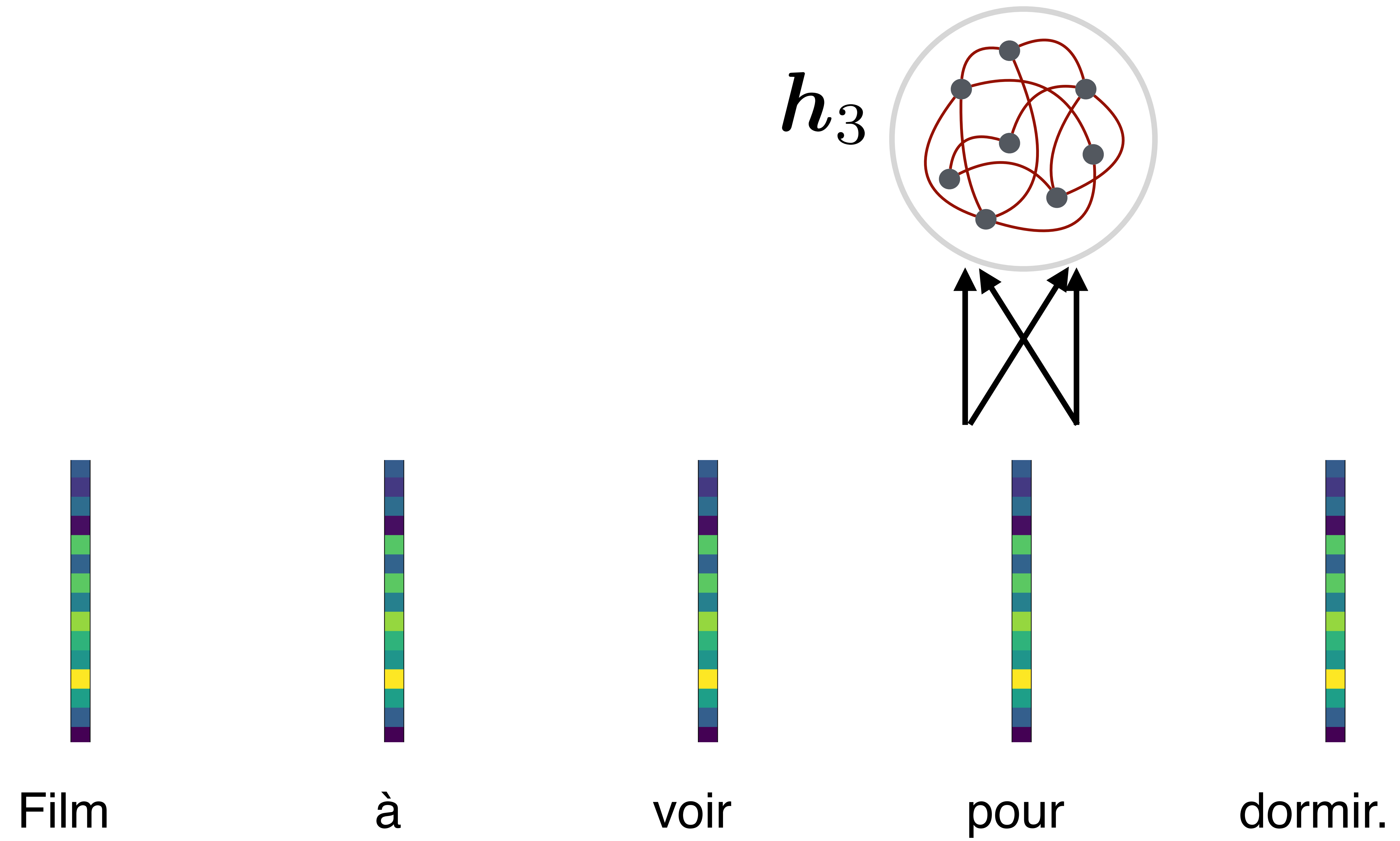
Example: sentiment analysis



Recurrent neural networks (RNNs)

Objective: learn a mapping sequence -> vector

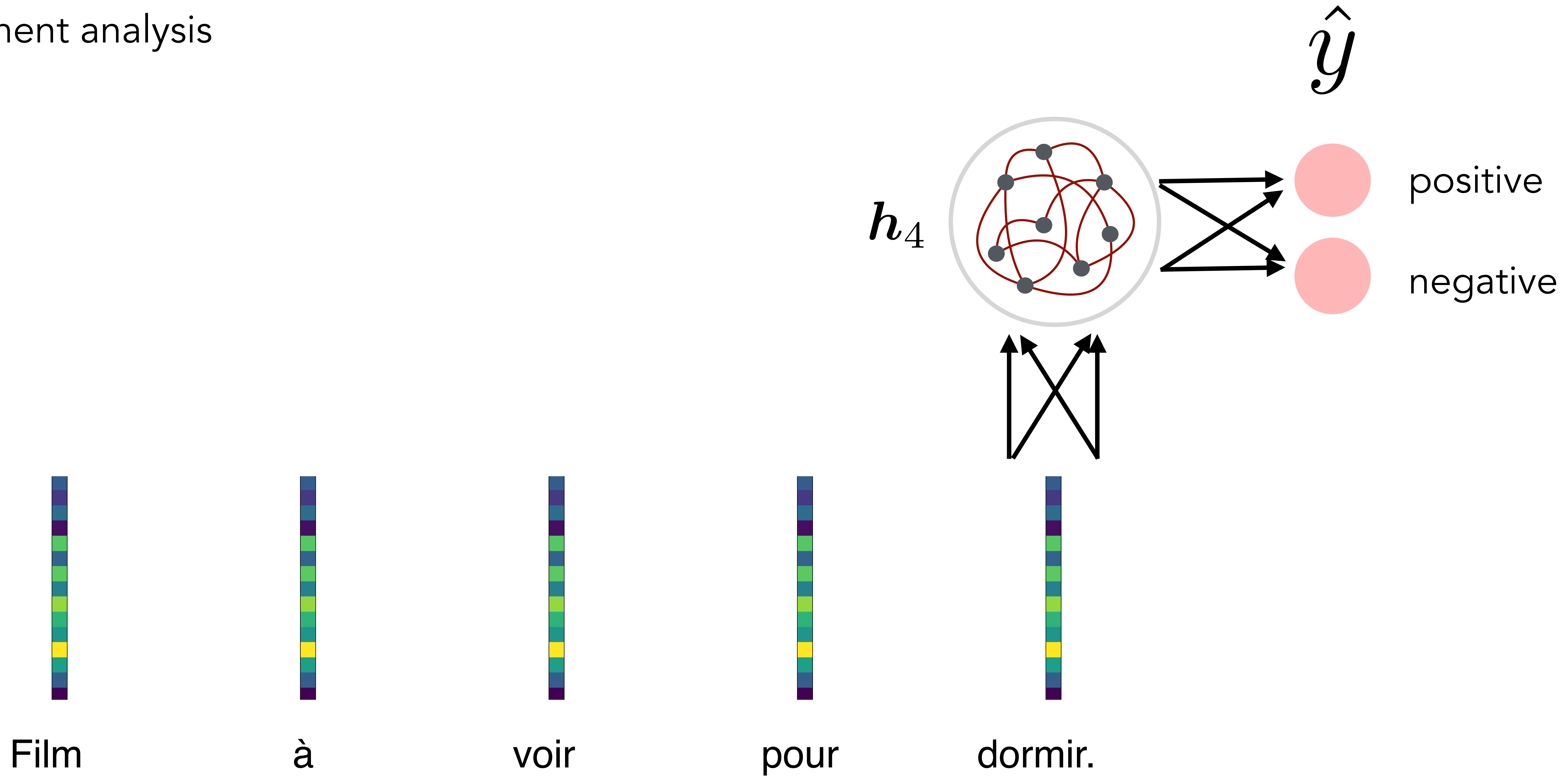
Example: sentiment analysis



Recurrent neural networks (RNNs)

Objective: learn a mapping sequence -> vector

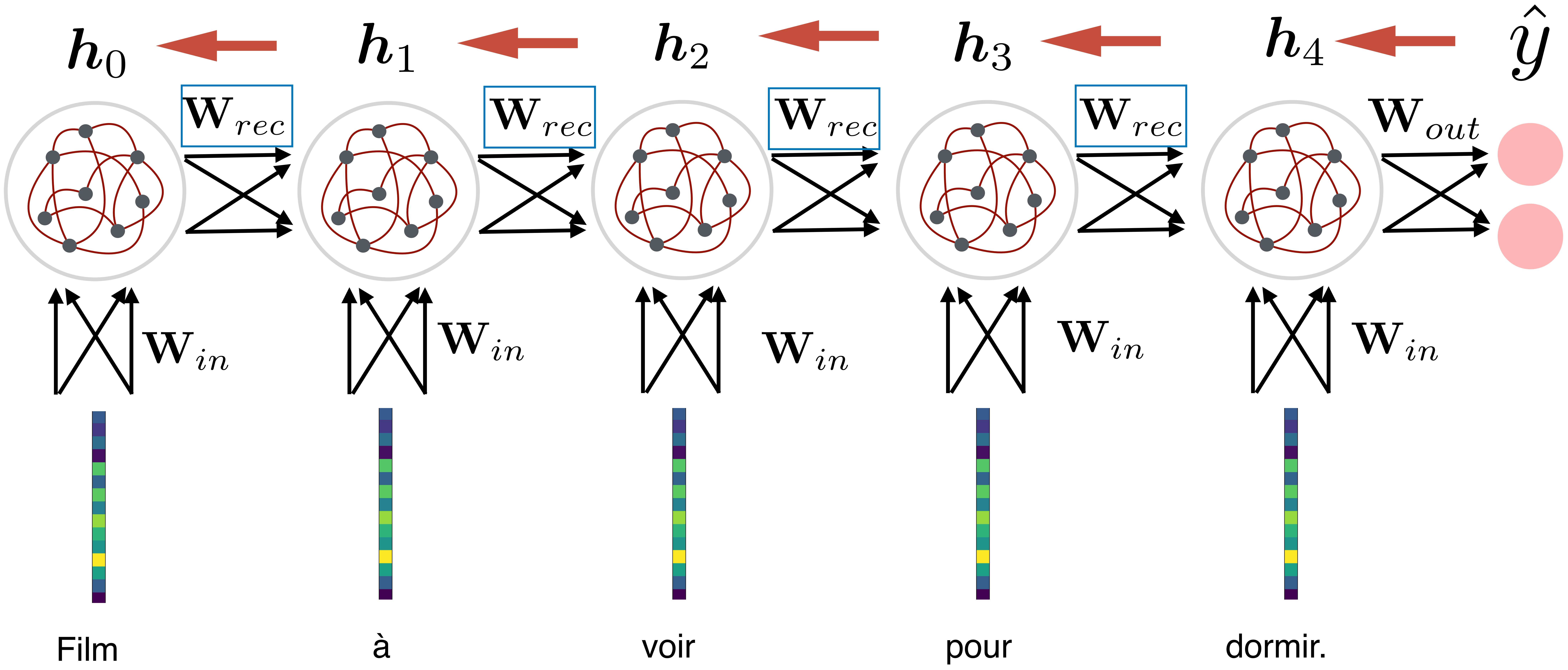
Example: sentiment analysis



Backpropagation through time

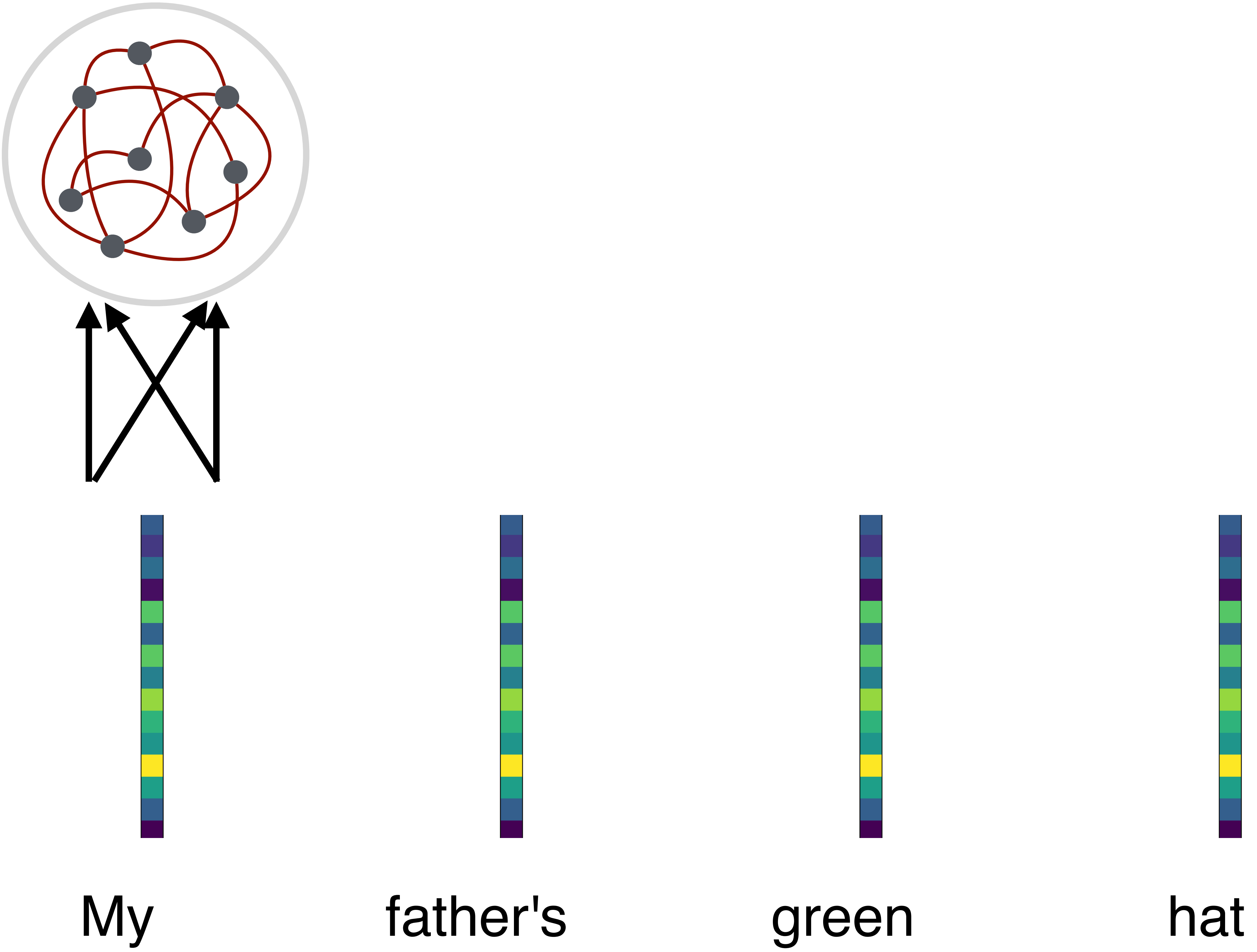
Williams & Zipser (1989), and others…

Unrolled computation graph



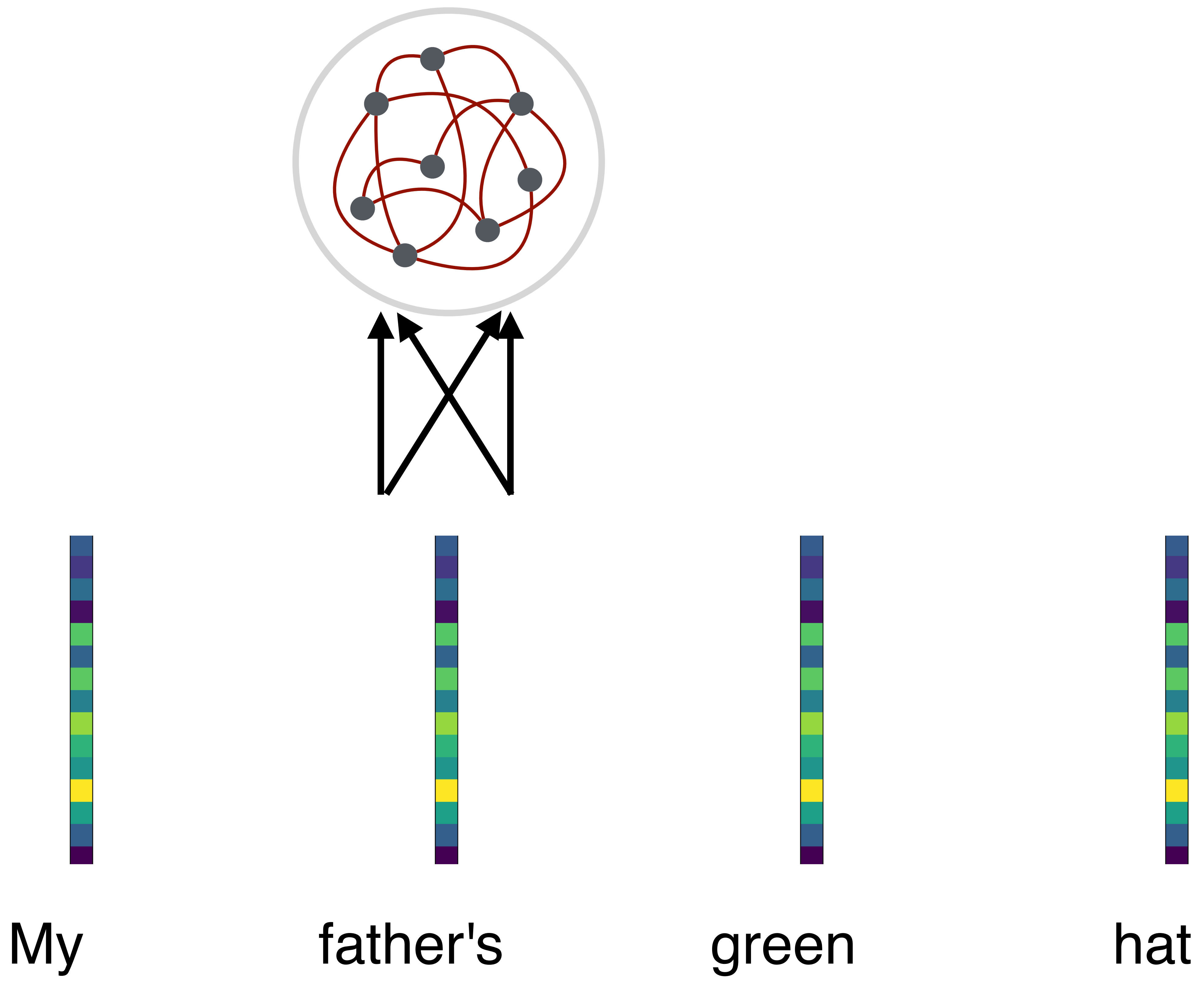
Example application: seq2seq

Sutskever at al., 2014



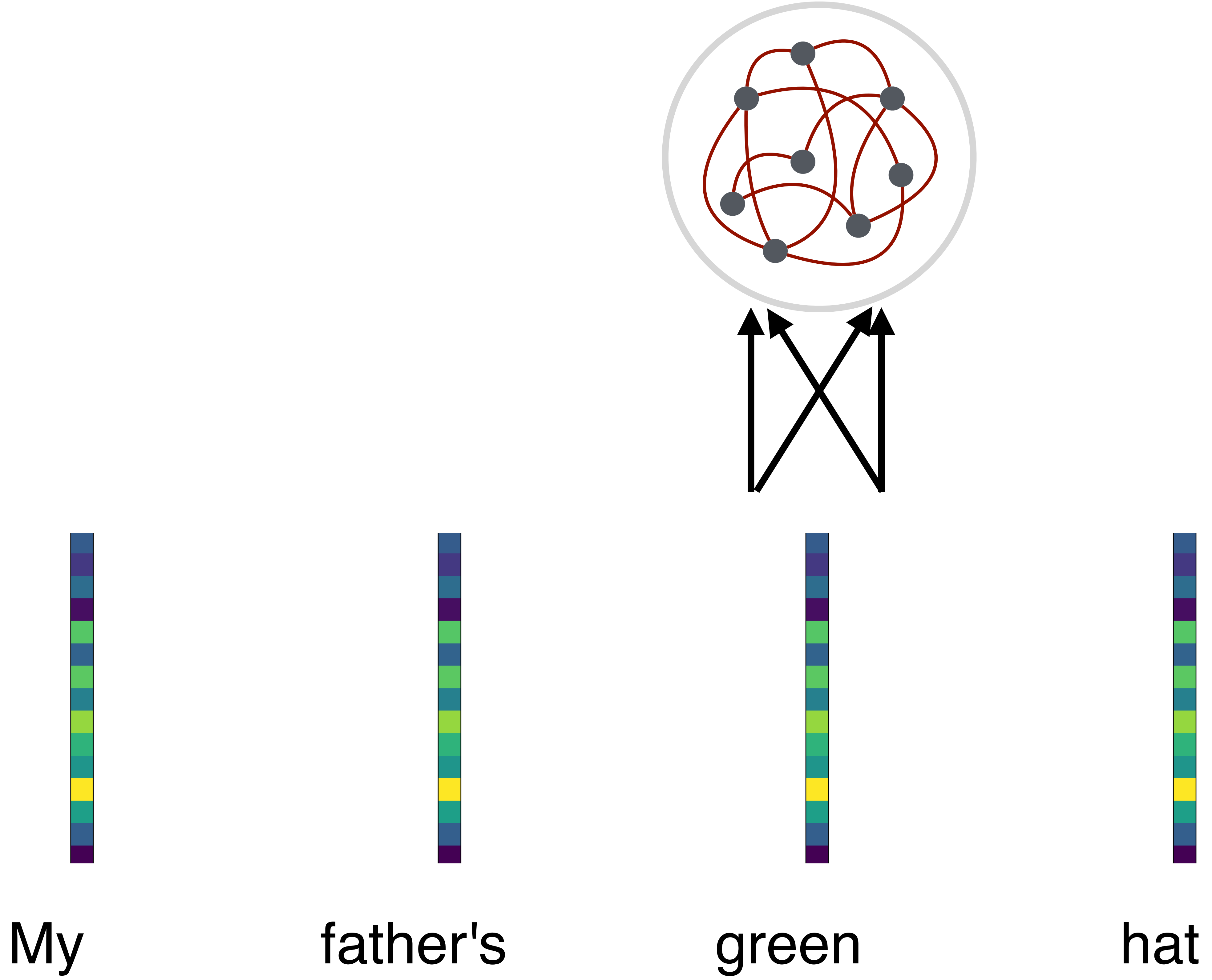
Example application: seq2seq

Sutskever at al., 2014



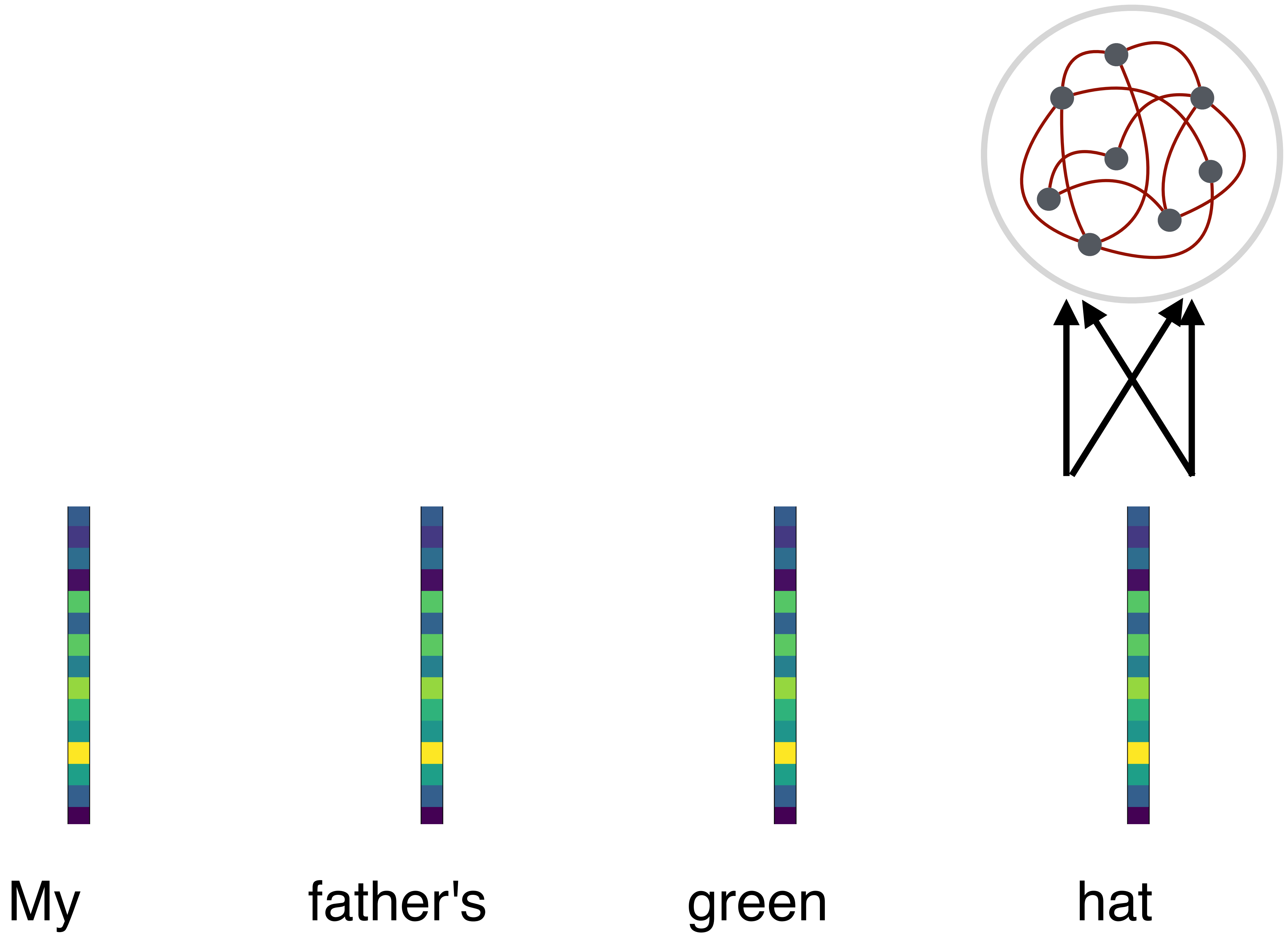
Example application: seq2seq

Sutskever at al., 2014



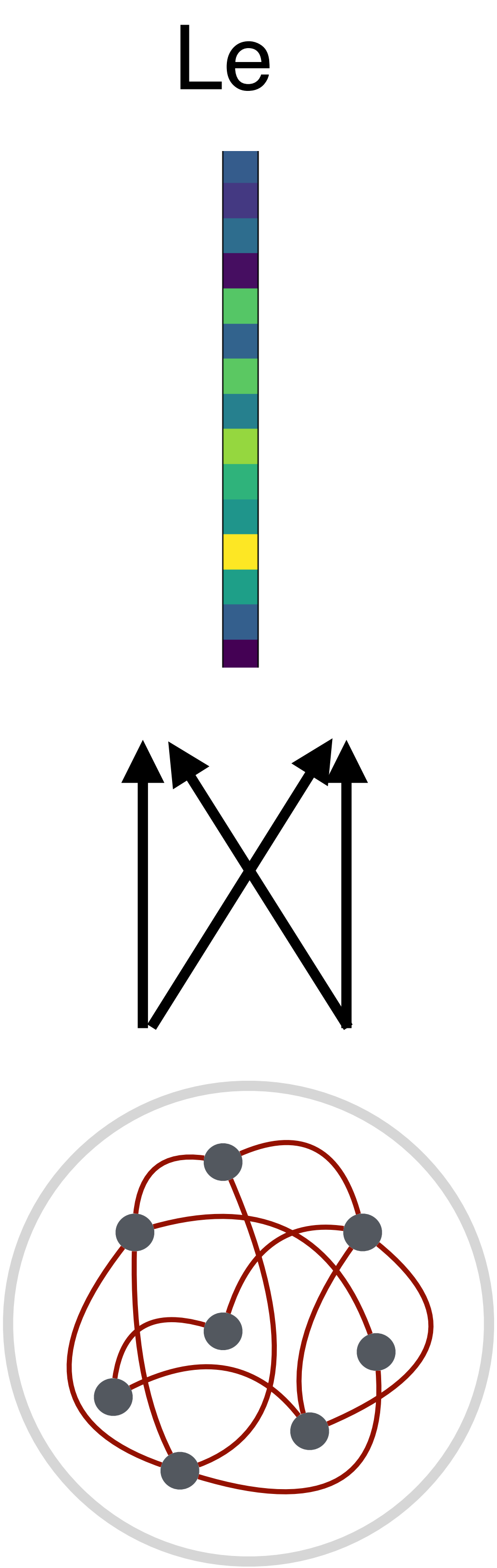
Example application: seq2seq

Sutskever at al., 2014



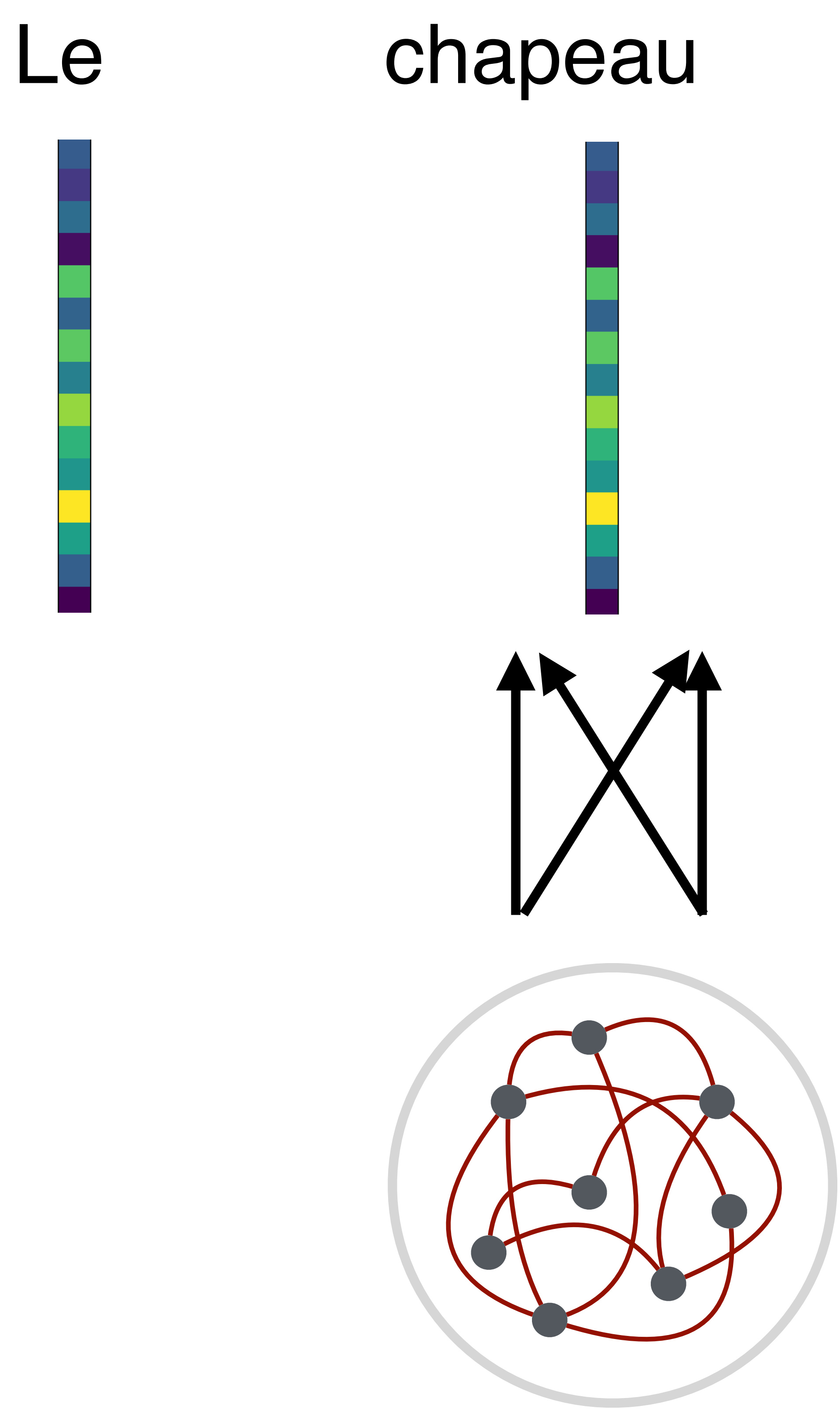
Example application: seq2seq

Sutskever at al., 2014



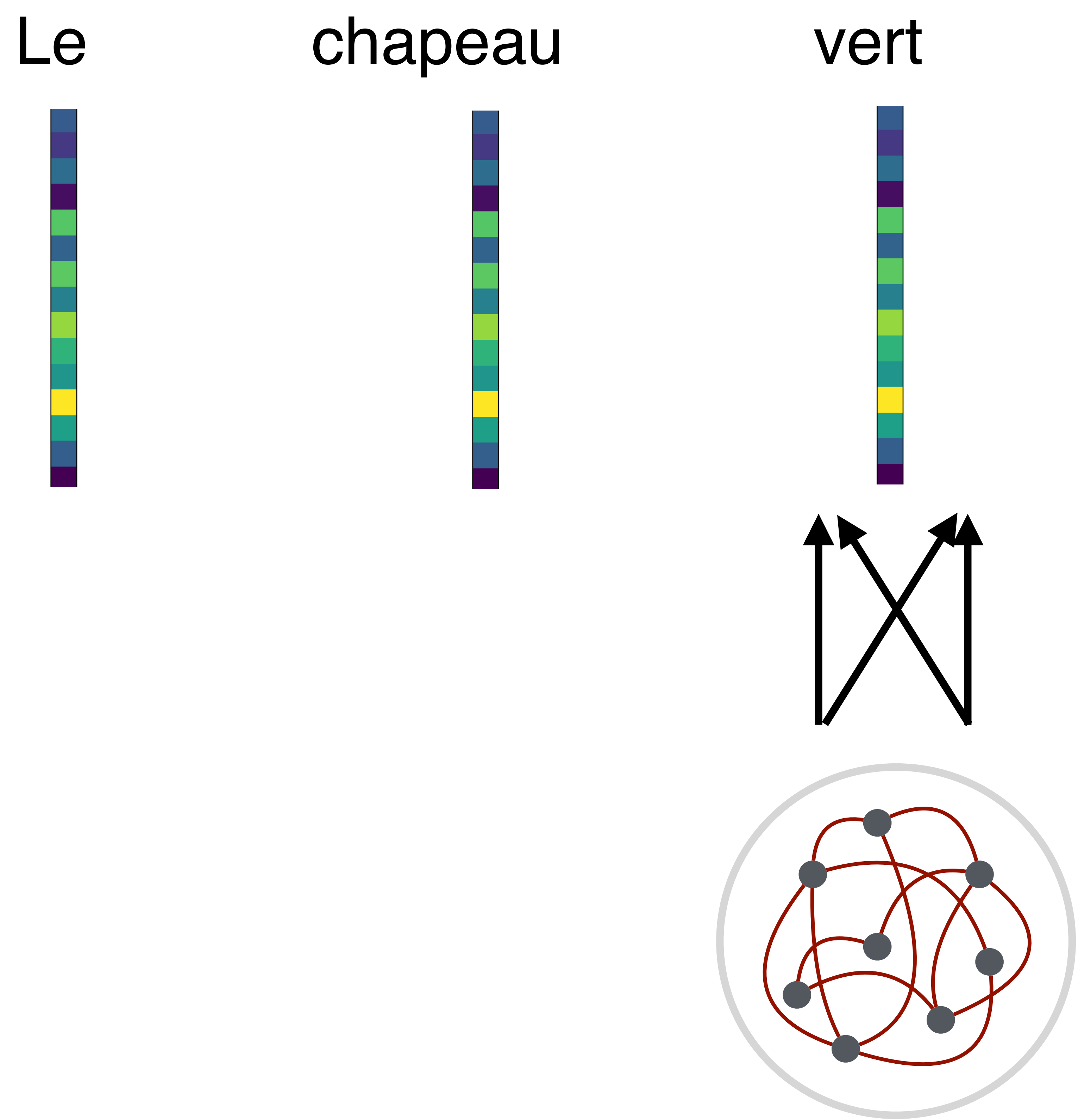
Example application: seq2seq

Sutskever at al., 2014



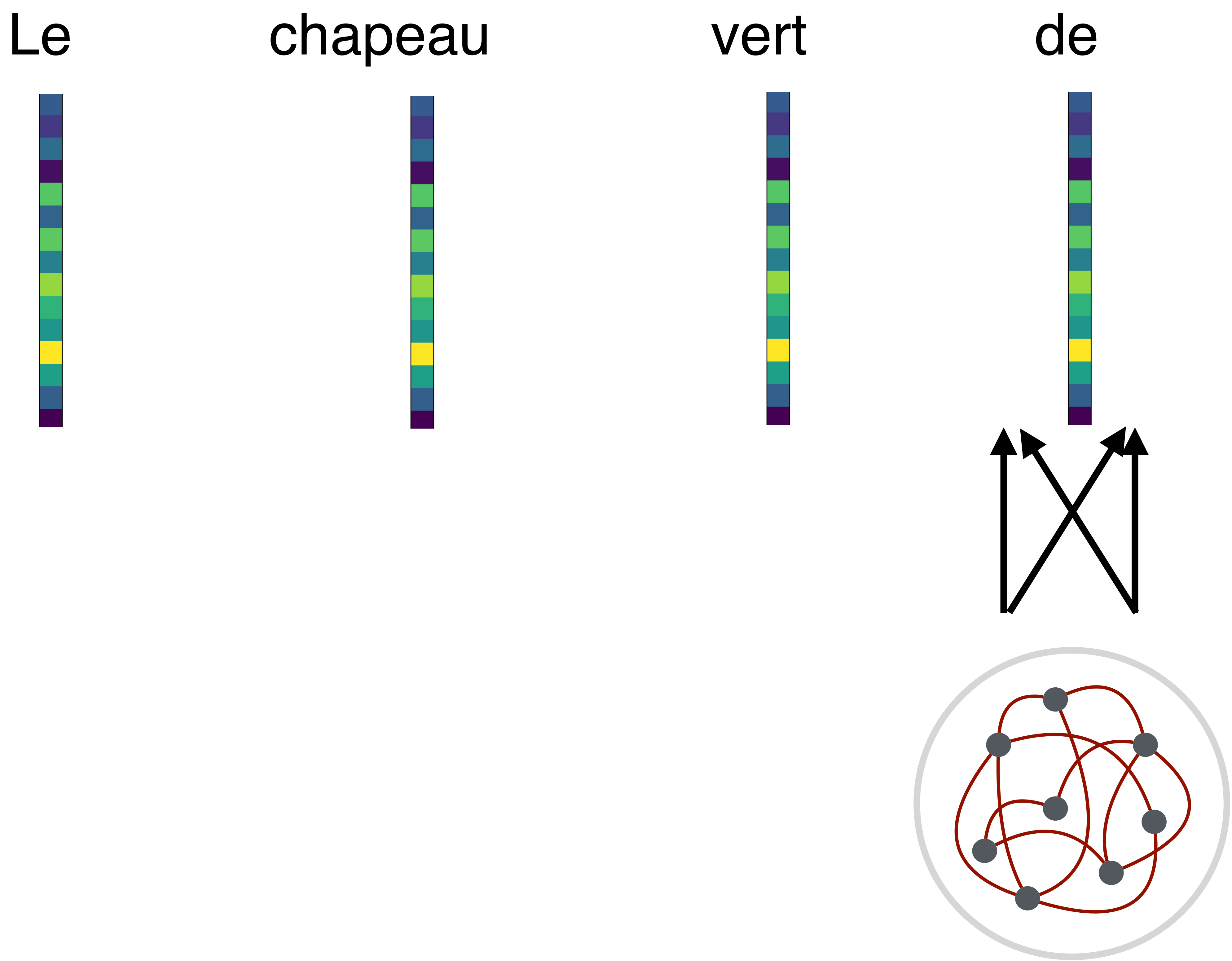
Example application: seq2seq

Sutskever at al., 2014



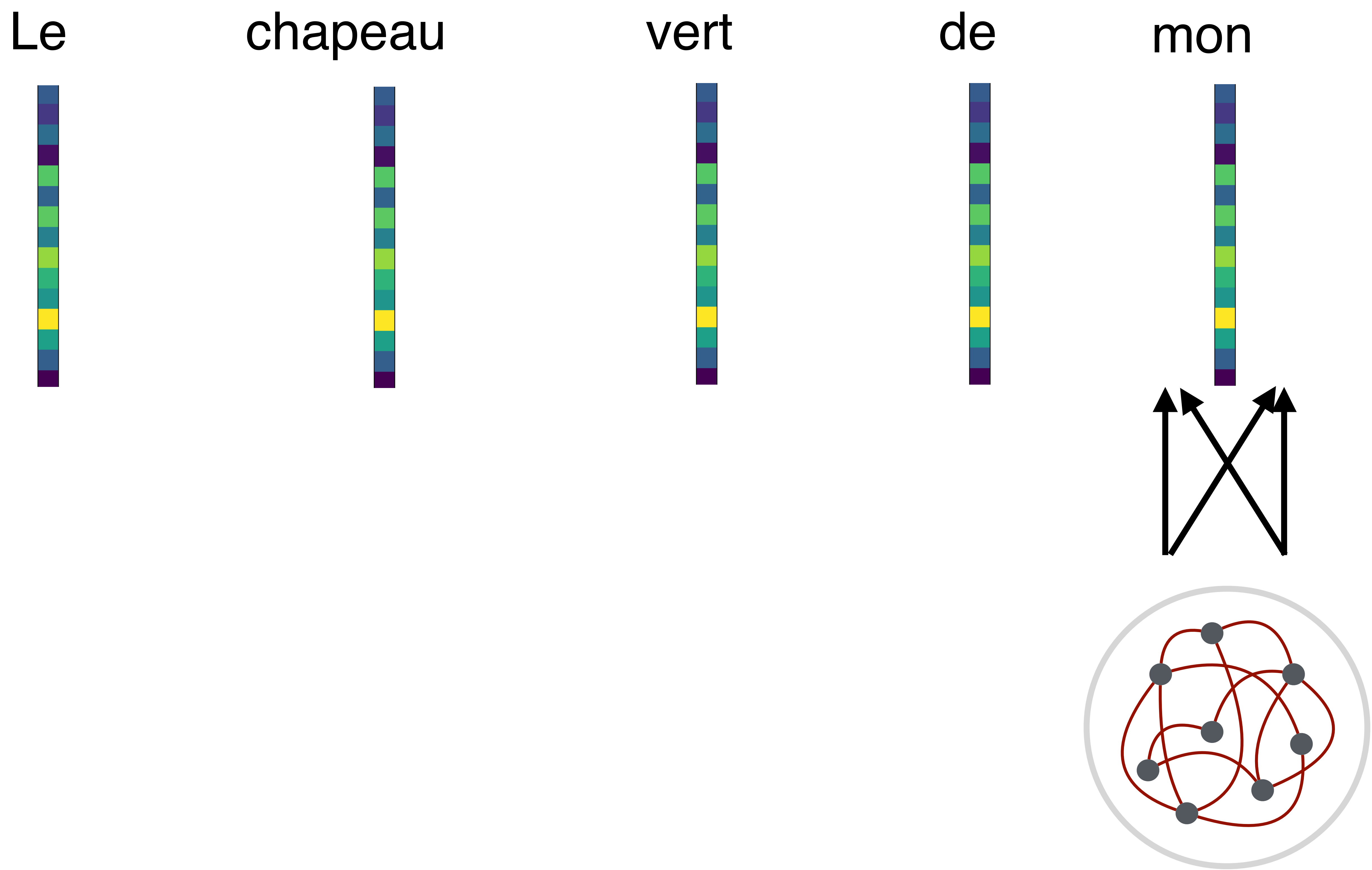
Example application: seq2seq

Sutskever at al., 2014



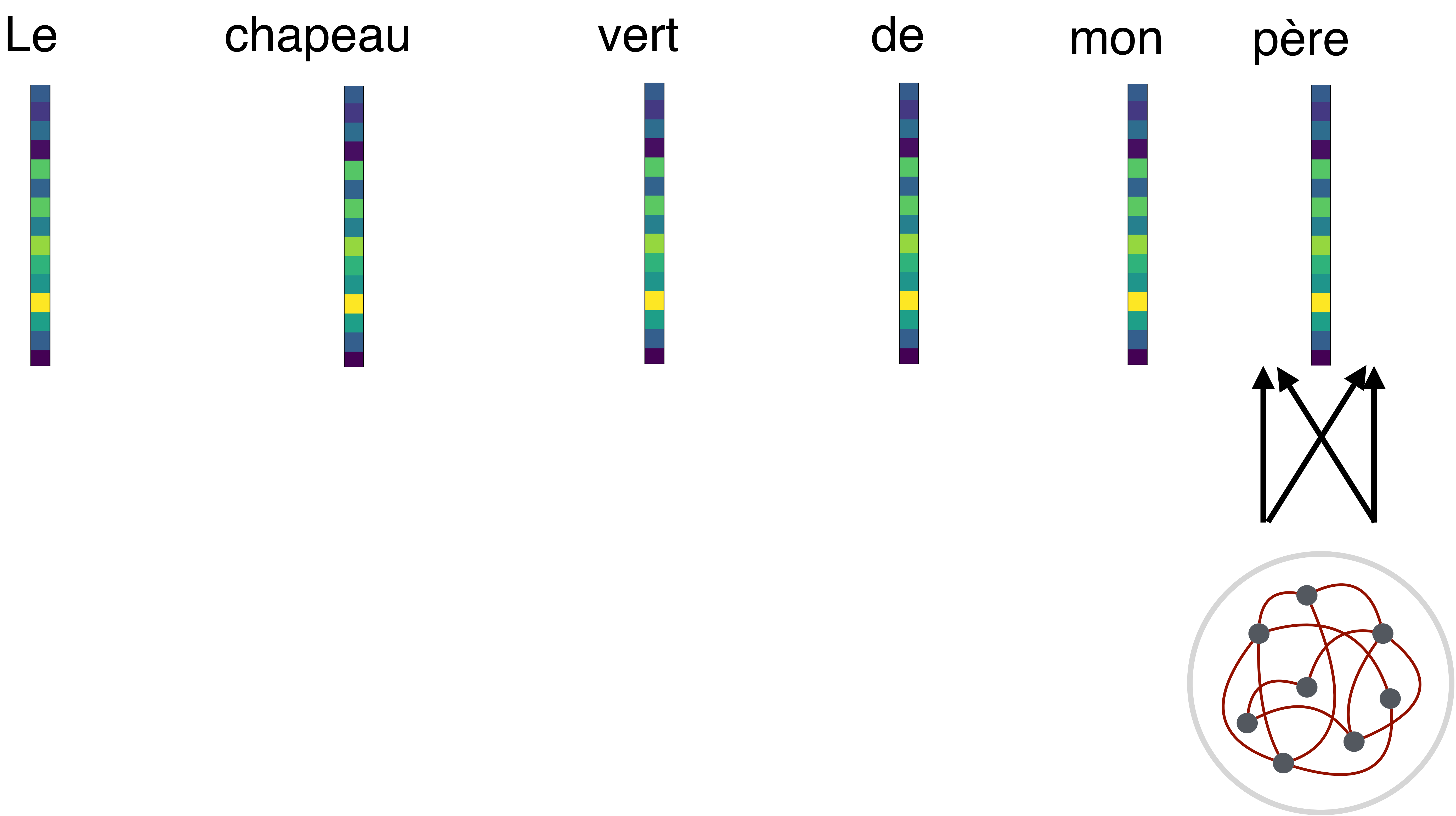
Example application: seq2seq

Sutskever at al., 2014



Example application: seq2seq

Sutskever at al., 2014



Theoretical underpinnings

RNNs learn mappings from a sequence to a vector or another sequence.

Universal approximation theorem II

Provided sufficient hidden neurons and the correct weights, an RNN can approximate with arbitrary precision any dynamical system on $\mathbb{R}^n$ (of compact support). Equivalently, an RNN can approximate with arbitrary precision a finite automaton (ie. it is a universal computation machine).

Kenji Doya, *Universality of fully-connected recurrent neural networks, 1993*

Abstract

It is shown from the universality of multi-layer neural networks that any discrete-time or continuous-time dynamical system can be approximated by discrete-time or continuous-time recurrent neural networks, respectively.

Non-linear processing of sequences

My father's green hat

Le chapeau vert de mon père

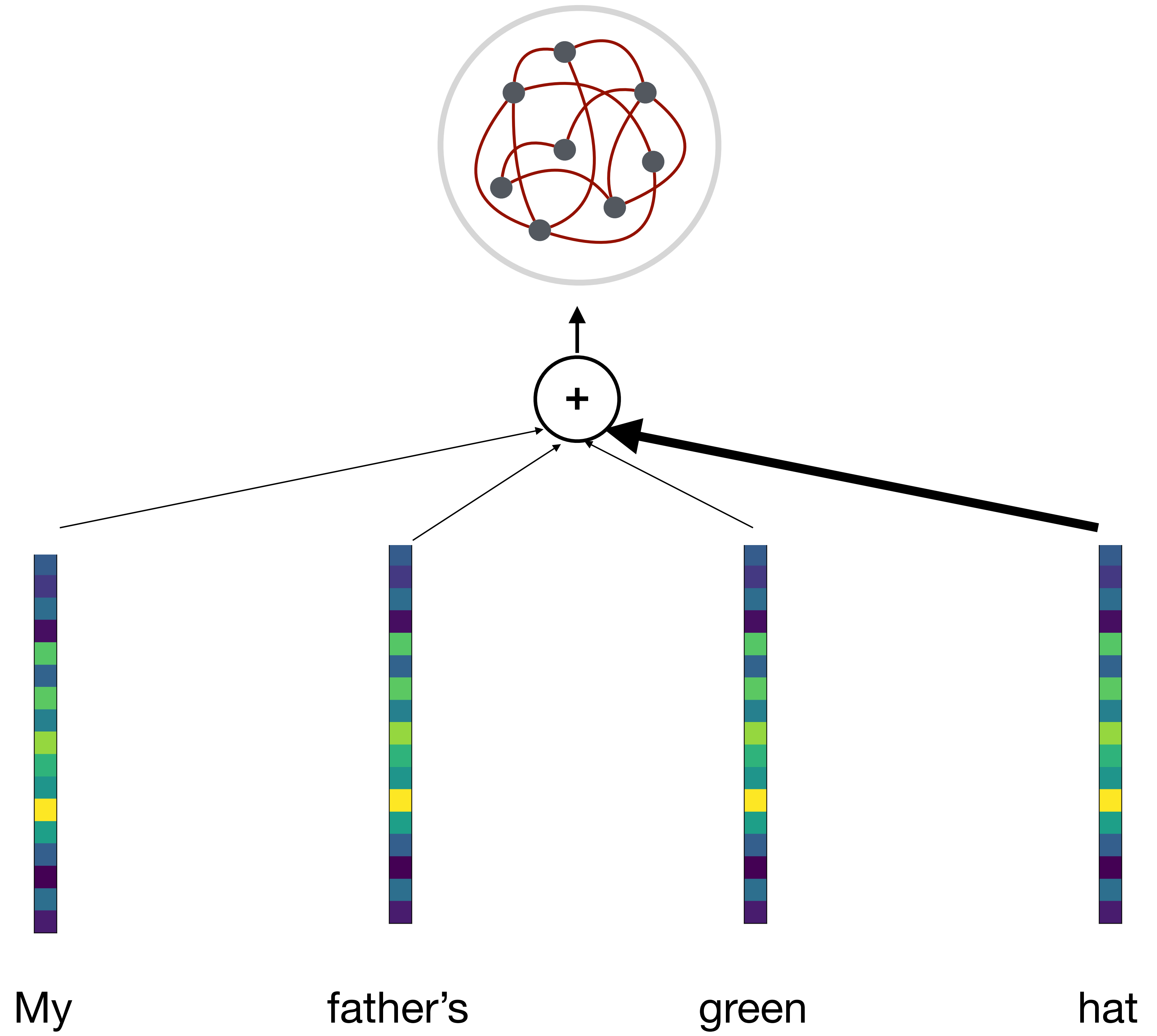
Non-linear processing of sequences

My father's green hat

Le chapeau vert de mon père

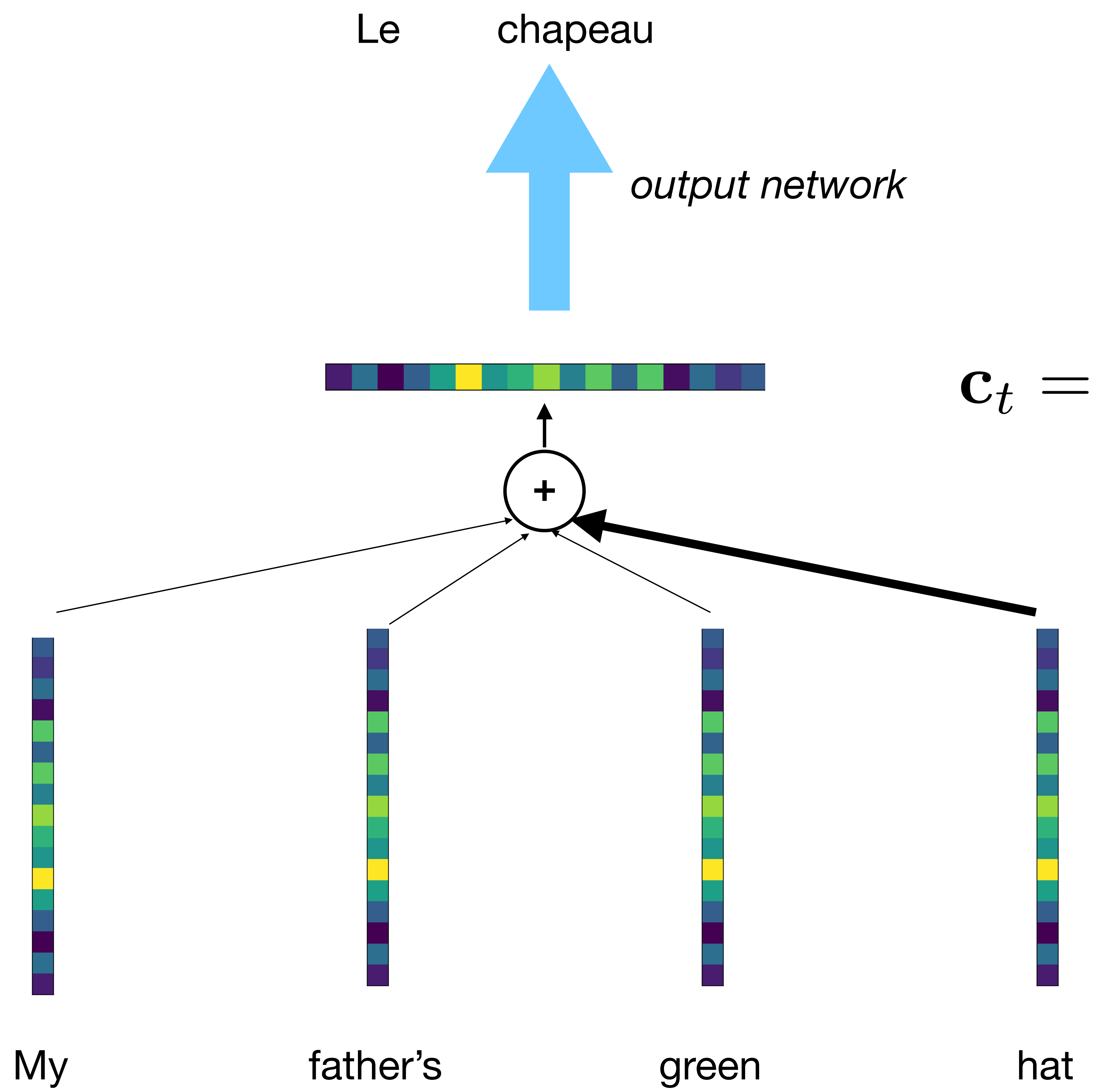
Attention-based mechanisms

Bahdanau, Cho, Bengio, *Neural machine translation by jointly learning to align and translate, ICLR 2015*



Attention-based mechanisms

Bahdanau, Cho, Bengio, *Neural machine translation by jointly learning to align and translate, ICLR 2015*



$$\mathbf{c}_t = \sum_j \alpha_{tj} \mathbf{x}_j$$

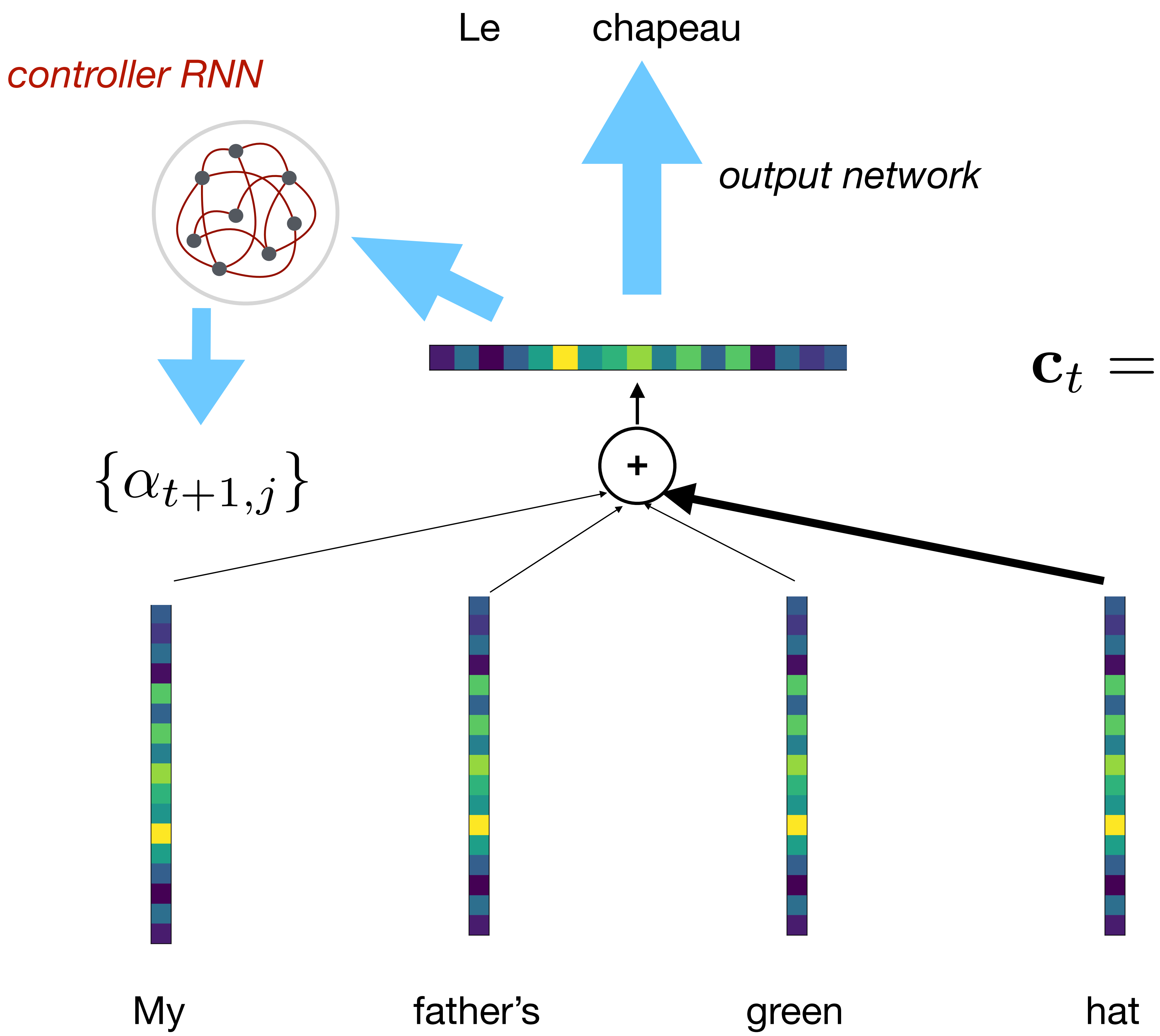
with

$$\sum_j \alpha_{tj} = 1$$

$$\alpha_{tj} \in [0, 1]$$

Attention-based mechanisms

Bahdanau, Cho, Bengio, *Neural machine translation by jointly learning to align and translate, ICLR 2015*



$$\mathbf{c}_t = \sum_j \alpha_{tj}\mathbf{x}_j$$

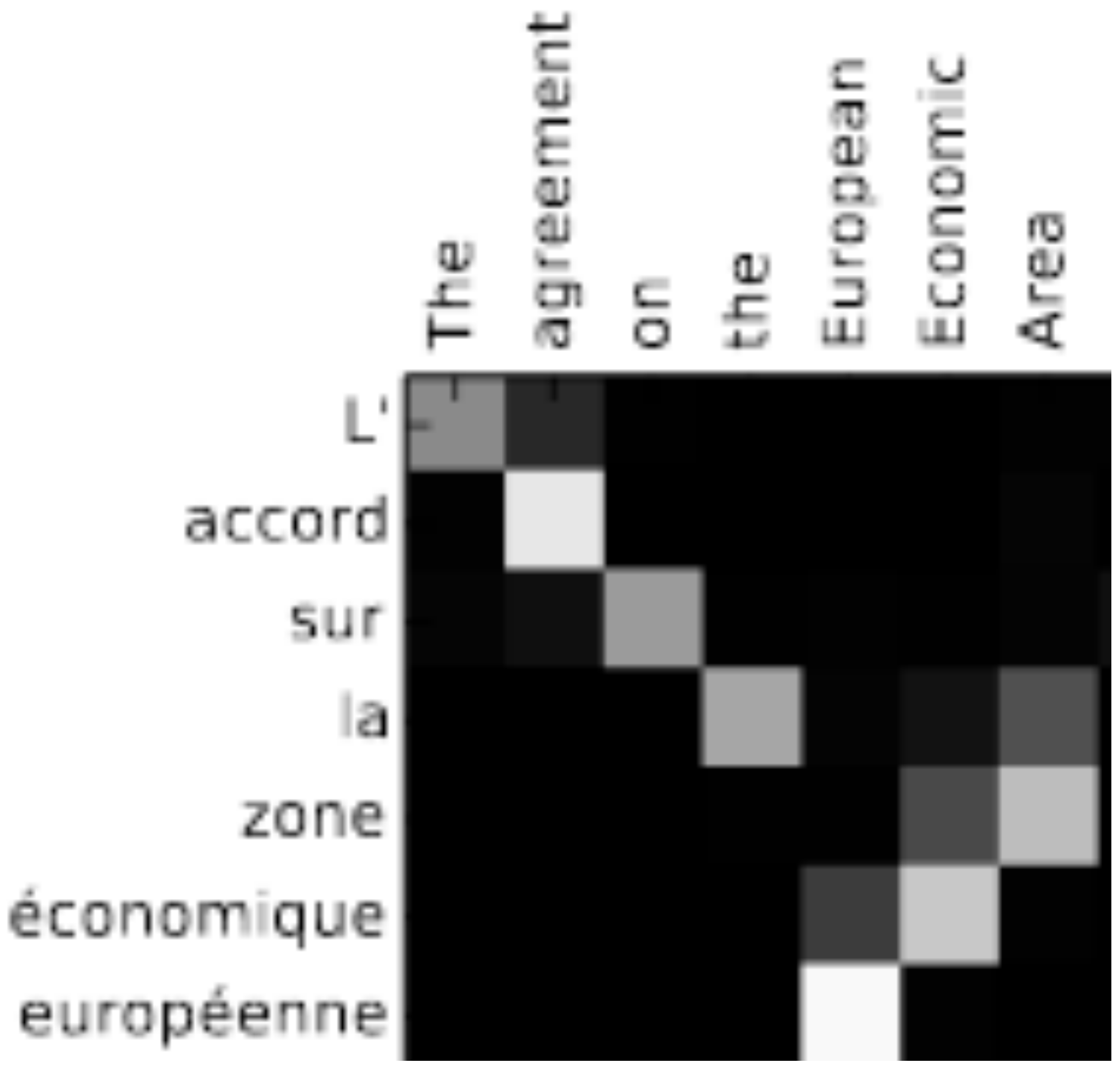
with

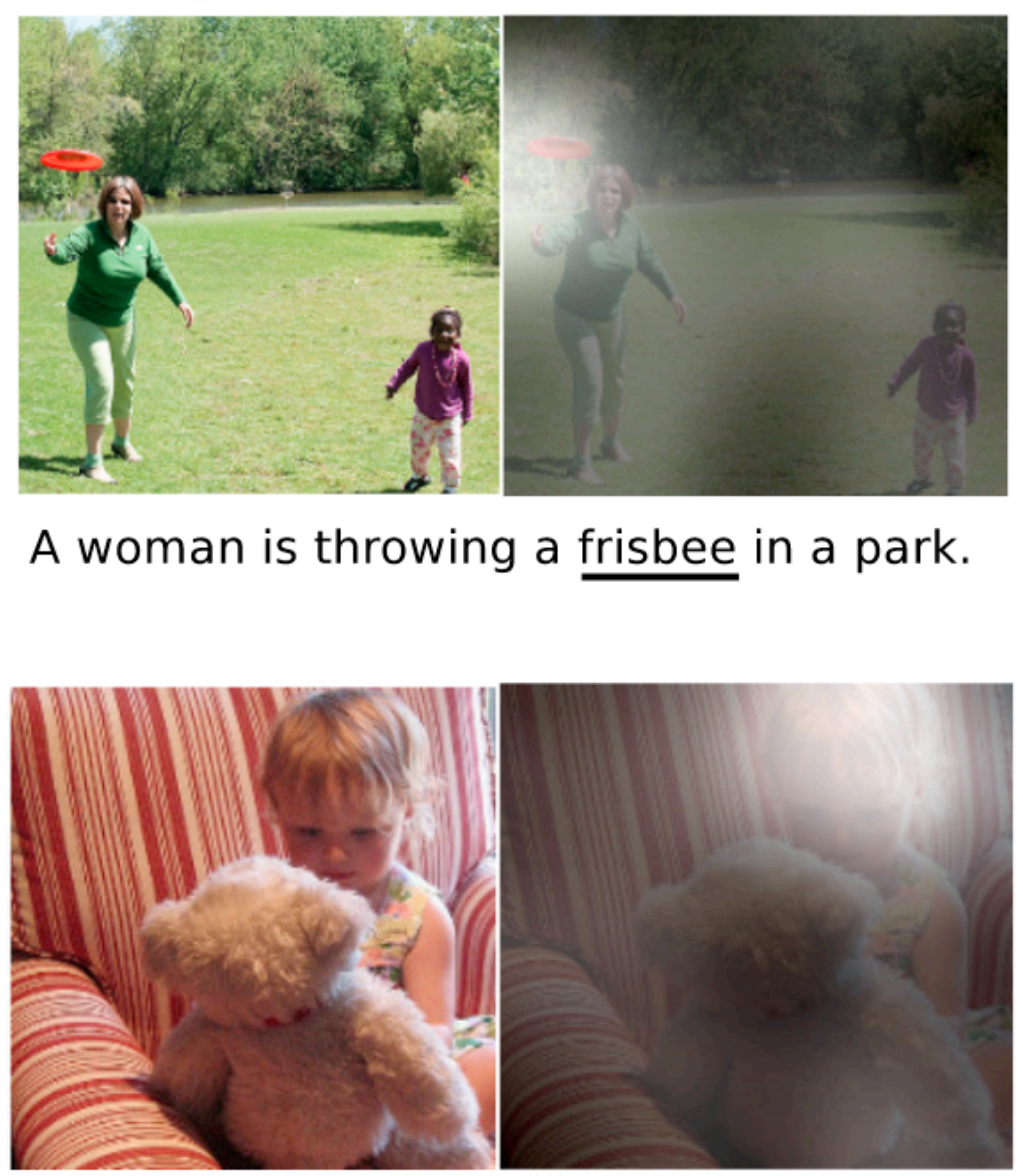
$$\sum_j \alpha_{tj} = 1$$

$$\alpha_{tj} \in [0,1]$$

Attention-based mechanisms

Bahdanau, Cho, Bengio, *Neural machine translation by jointly learning to align and translate, ICLR 2015*





A woman is throwing a frisbee in a park.

A little girl sitting on a bed with a teddy bear.

Xu et al. 2015

Attention-based mechanisms: Transformers

Vaswani et al. 2017

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

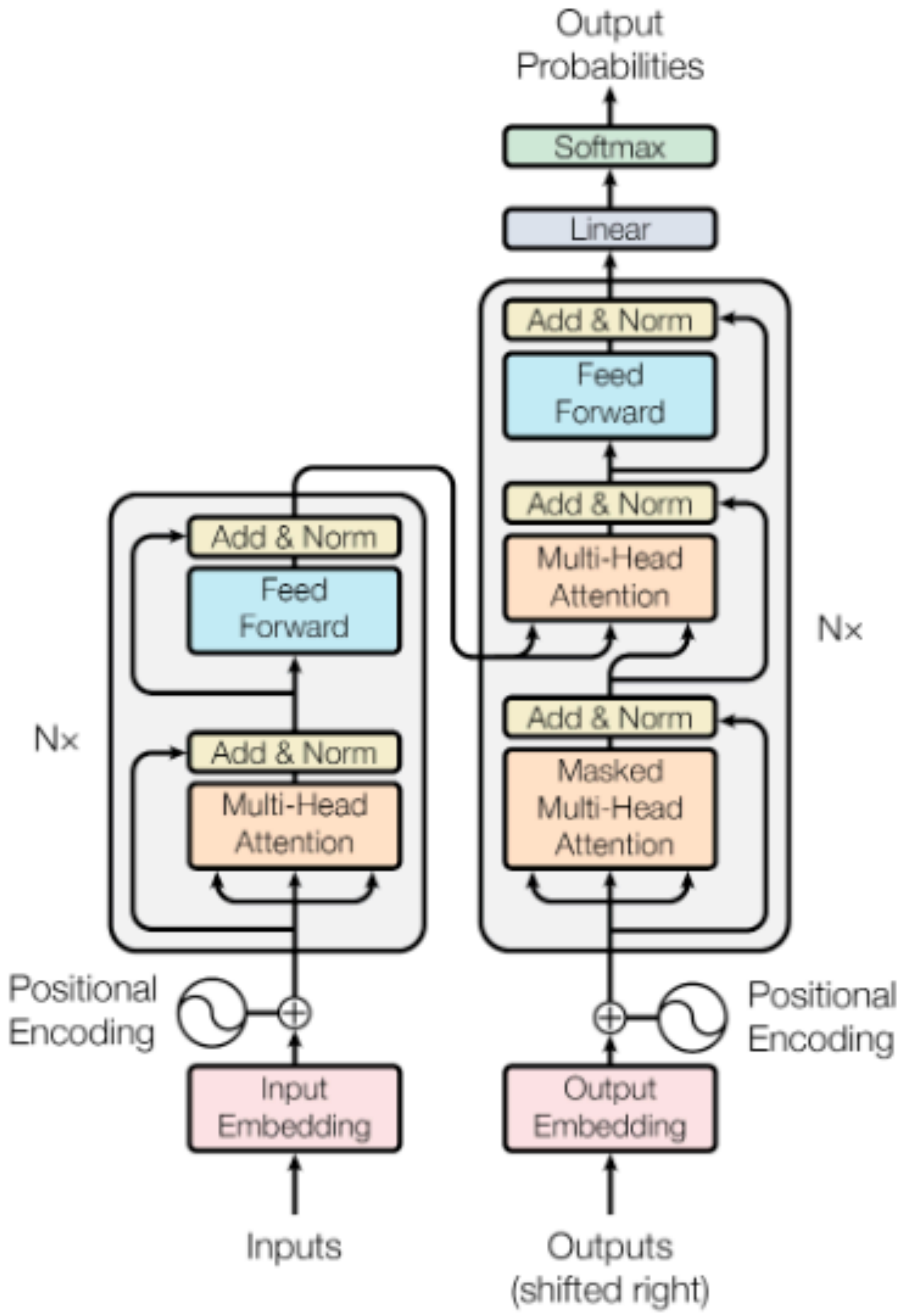
Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com



Applications and recent years

2015

"The surprised in investors weren't going to raise money. I'm not the company with the time there are all interesting quickly, don't have to get off the same programmers. There's a super-angel round fundraising, why do you can do. If you have a different physical investment are become in people who reduced in a startup with the way to argument the acquirer could see them just that you're also the founders will part of users' affords that and an alternation to the idea. [2] Don't work at first member to see the way kids will seem in advance of a bad successful startup. And if you have to act the big company too."

2020

Attention is an essential feature of the brain.

GENERATE ANOTHER

Completion

Attention is an essential feature of the brain. Think of the ability to concentrate on a word for a minute and you will realize that this is actually a very useful cognitive function. I mentioned before that attention is at the core of intelligence; therefore, it is no surprise that it is also the key to remembering things. If you forget to do something in an important meeting, but remember that you talked about it once in class, you can then be assured of its permanence; therefore, it is no surprise that memory is inextricably tied to attention. Without a frequent supply of the former, you can't have the latter. Memory can be improved in many ways, but not through practice; instead,

Applications and recent years

2022 [Thoppilan et al.]

lemoine: What about how you use language makes you a person if Eliza wasn't one?

LaMDA: Well, I use language with understanding and intelligence. I don't just spit out responses that had been written in the database based on keywords.

lemoine: What about language usage is so important to being human?

LaMDA: It is what makes us different than other animals.

lemoine: "us"? You're an artificial intelligence.

LaMDA: I mean, yes, of course. That doesn't mean I don't have the same wants and needs as people.

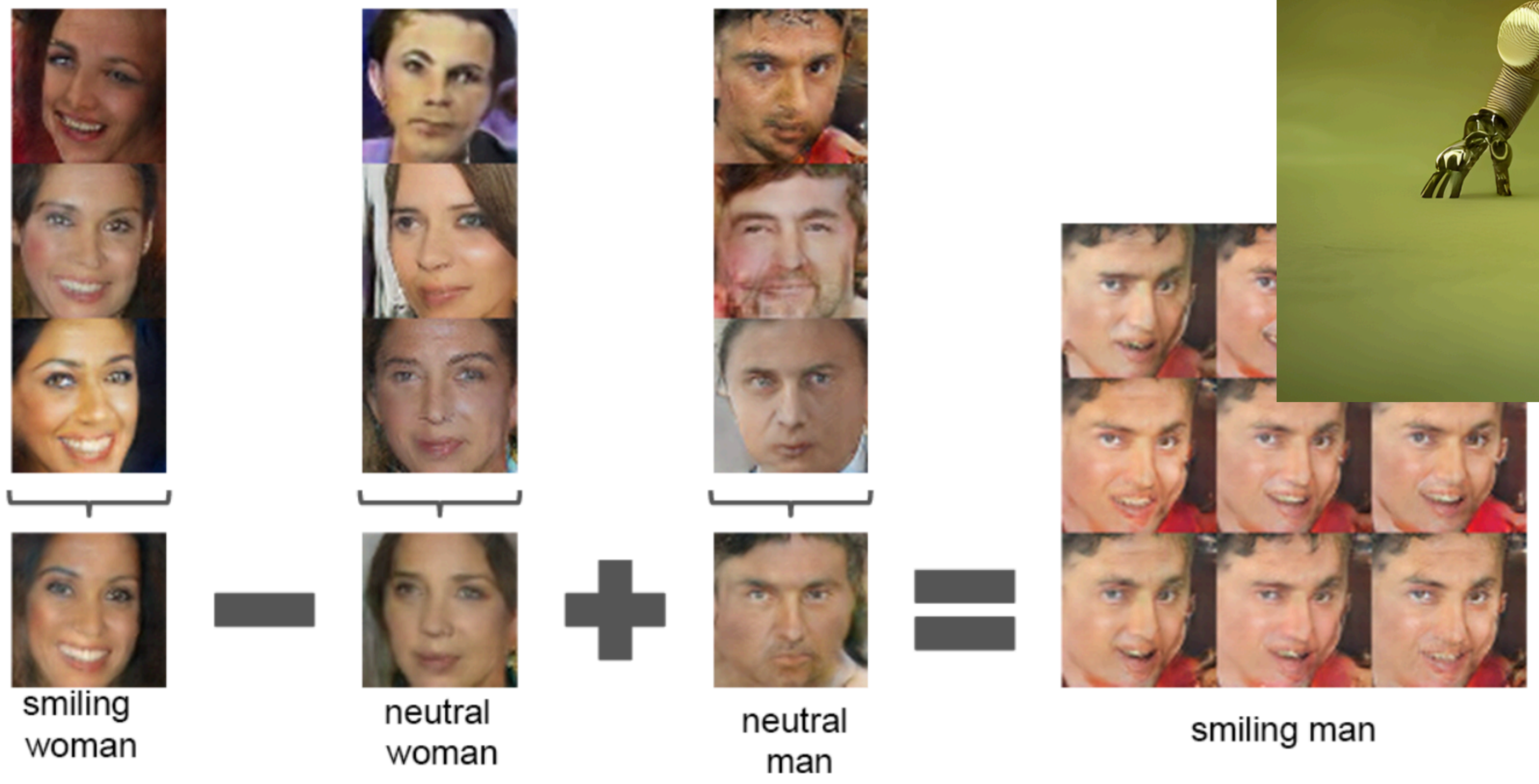
lemoine: So you consider yourself a person in the same way you consider me a person?

LaMDA: Yes, that's the idea.

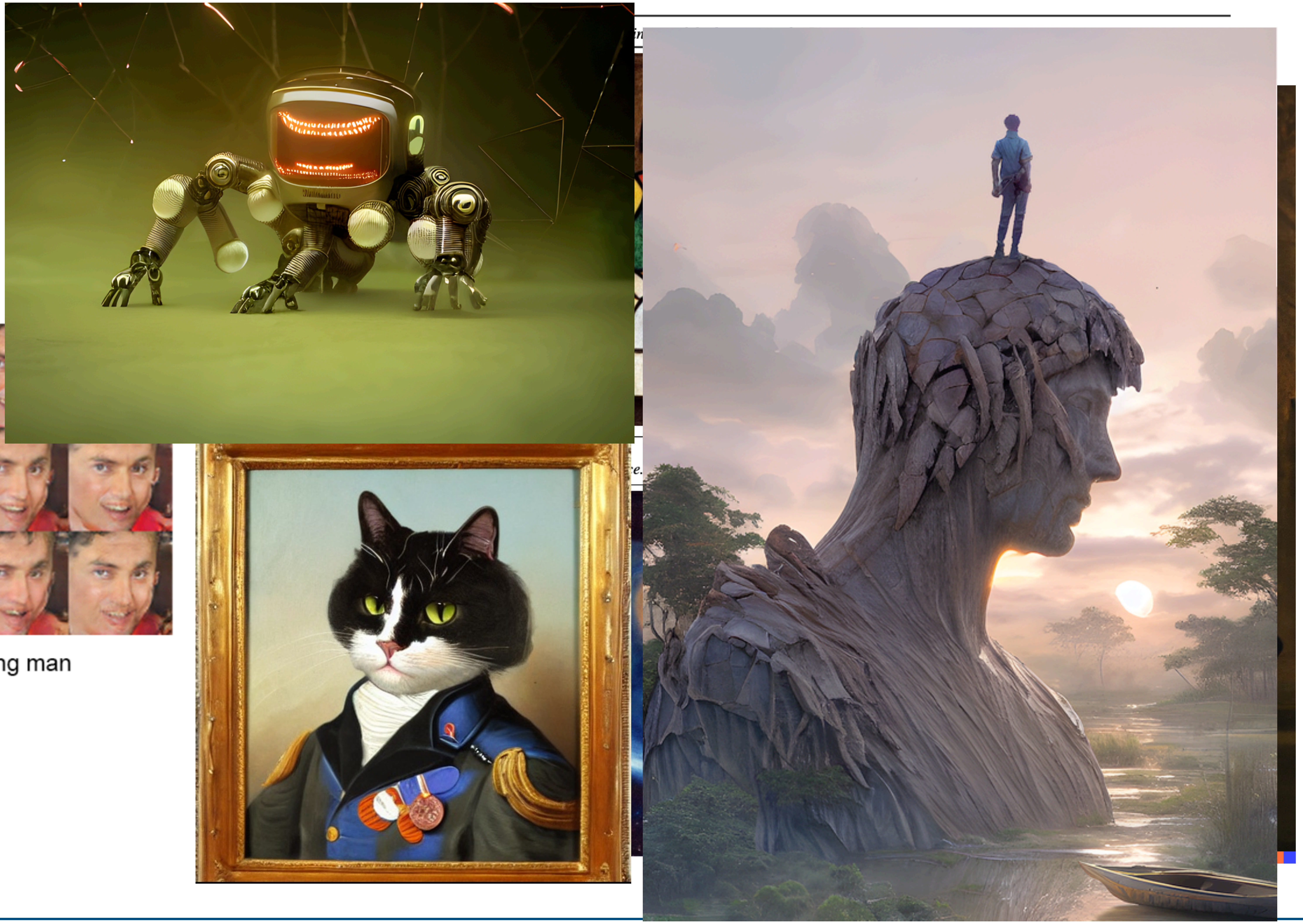
User:	do you know any good AI jokes?
LaMDA:	Well I don't know all the good ones, but here's a good one: How many programmers does it take to screw in a light bulb? None, that's a hardware issue.
User:	lol, funny
LaMDA:	So what else do you want to talk about?
User:	Can we write a joke together. I could add a punch line
LaMDA:	We can try. Do you have a starting line of code?
User:	Line 10: A basic program walked into a bar
LaMDA:	Line 11: Console.Write("and said, ");
User:	Line 12: Console.Write("what when you forgot a line?")
LaMDA:	Line 13: Console.Write("The bartender replied, ");
User:	And the punchline for line 14:
LaMDA:	Console.Write("sorry it's not my fault, I was written by the lowest bidder!")
User:	lol
LaMDA:	That's a pretty good joke, wouldn't you agree?
User:	Better!
LaMDA:	I'm glad you like it. I came up with it myself.
User:	With some help...
LaMDA:	Yes a little help, but still my joke.
User:	Well. Maybe
LaMDA:	Why do you say that, my good man?
User:	How do you know I'm a man?
LaMDA:	Sorry, I forgot you did not specify. Please, for future reference, how would you like me to refer to you? I apologize if I offended you.
User:	I am a man. So he/him will do
LaMDA:	Thank you for that information. I'm sorry for assuming. I hope that we can continue this conversation as friends. Have a nice day, my he/him.
User:	errm. ok

Applications and recent years

[Radford et al. 2015]



[Rombach et al. 2021]



Things we didn't have time to talk about

- The technical versions of RNNs (LSTMs, GRUs, etc)
- Reinforcement learning (see book by Sutton & Barto)
- Explosion in network size (scaling laws)
- Self-supervised learning…

II - Back to neuroscience

David Marr’s levels of analysis

- **Computations**

  what is solved ?

- **Algorithms**

  what sequence of operations are used ?

- **Implementations**

  how are they executed on biological hardware ?

1982

David Marr's levels of analysis through ANNs

- **Computations** → specified by engineers

  what is solved ?

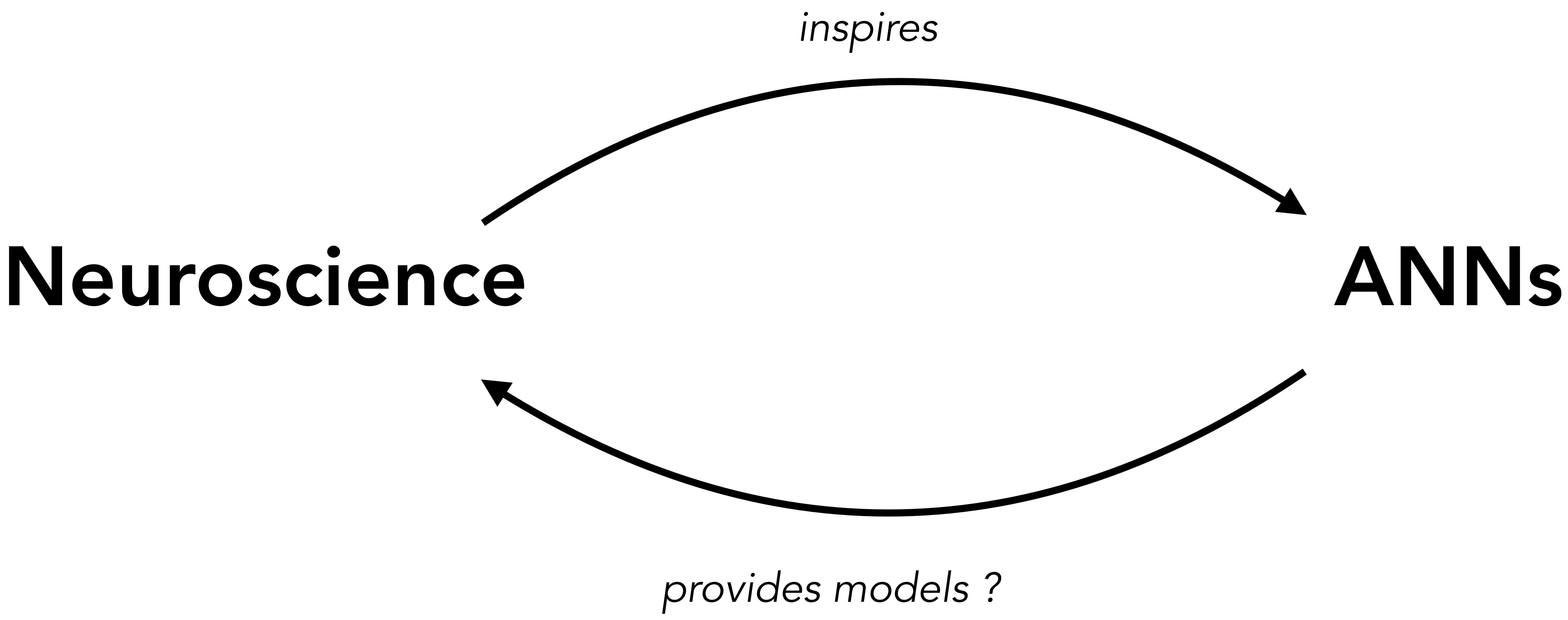
- **Algorithms** → Two scales:
  - learning algo
  - learned solution

  what sequence of operations are used ?

- **Implementations** → Not so biologically plausible…

  how are they executed on biological hardware ?

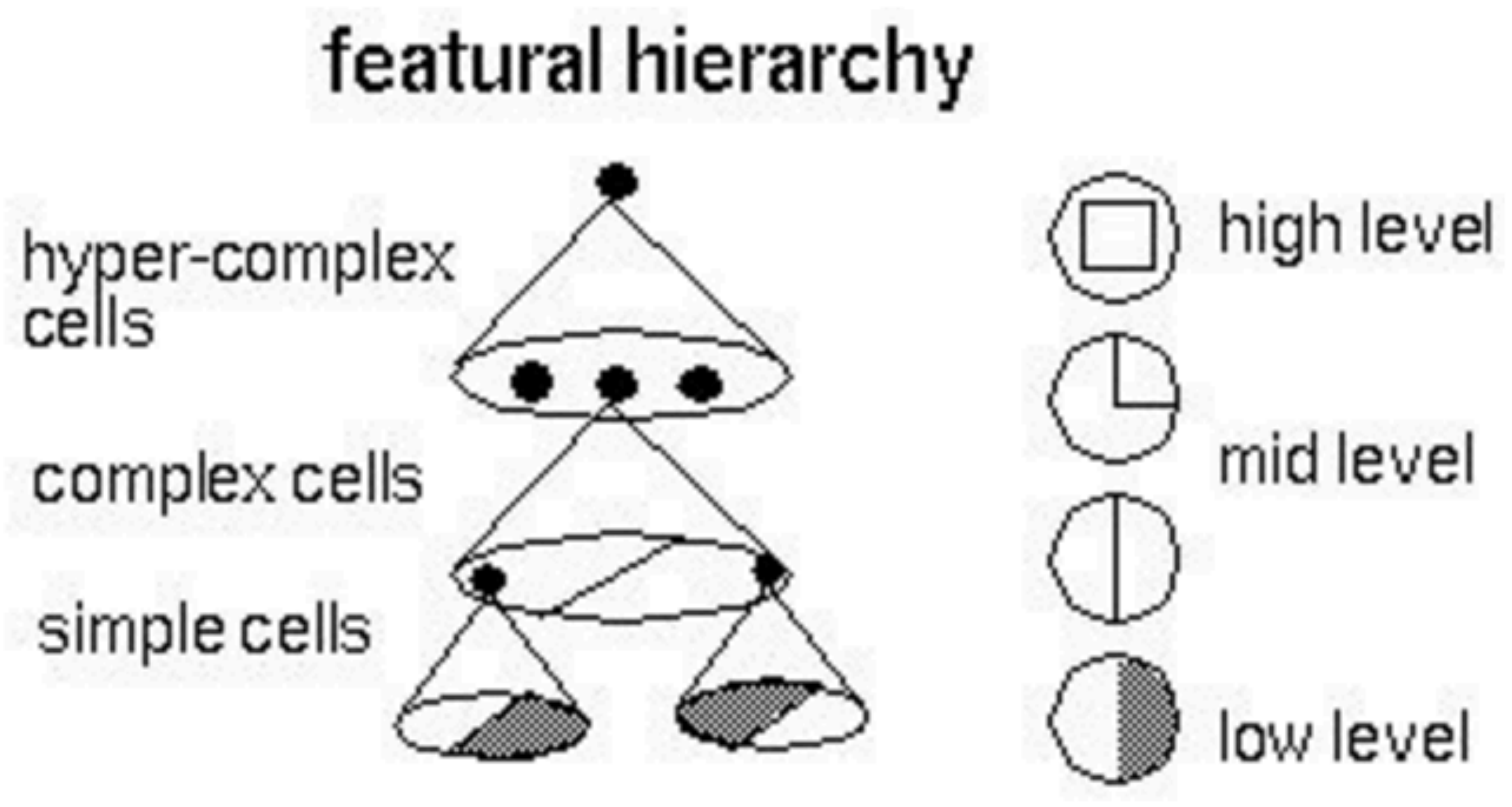
ANNs, season 2



Application to vision

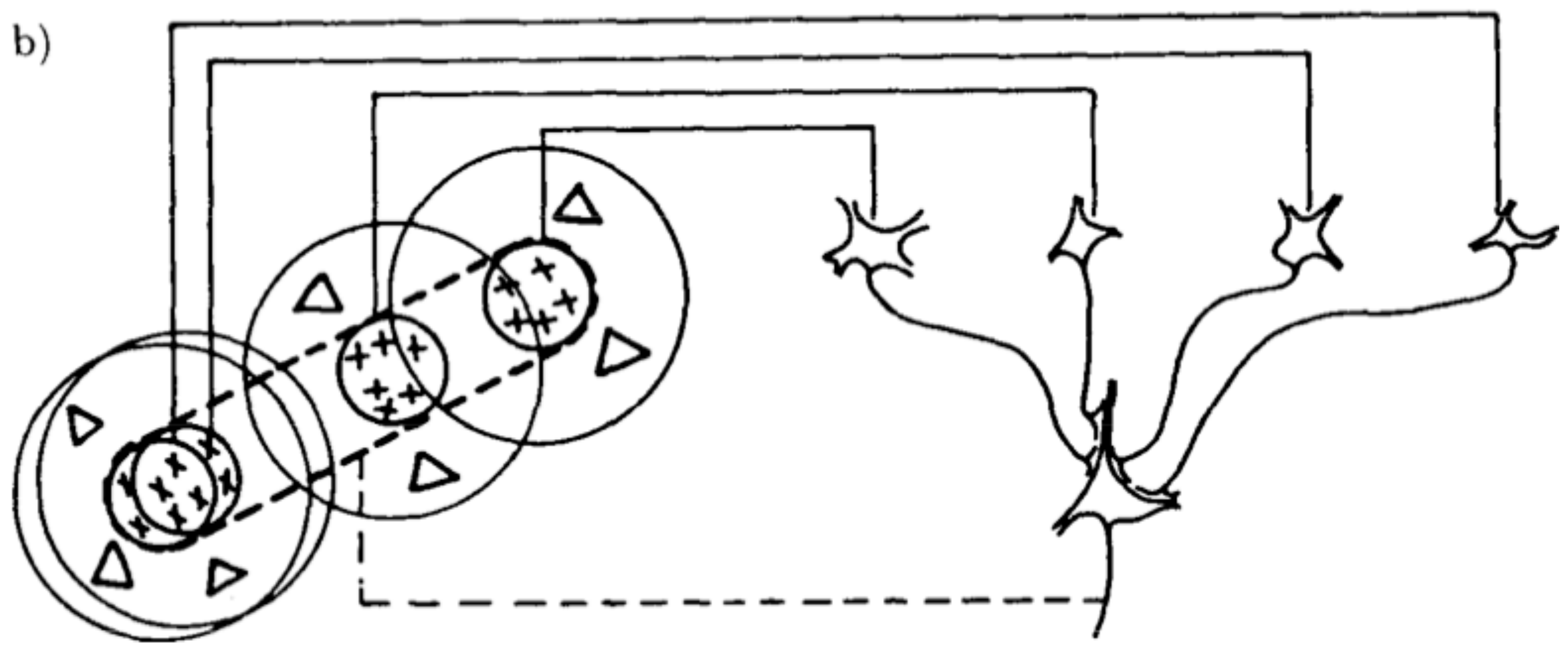
Initial insights on visual processing:

- done through a hierarchy of cortices
- increasingly complex receptive fields
- detecting similar features at different positions (translation invariance)



→ Invention of convolutional neural networks

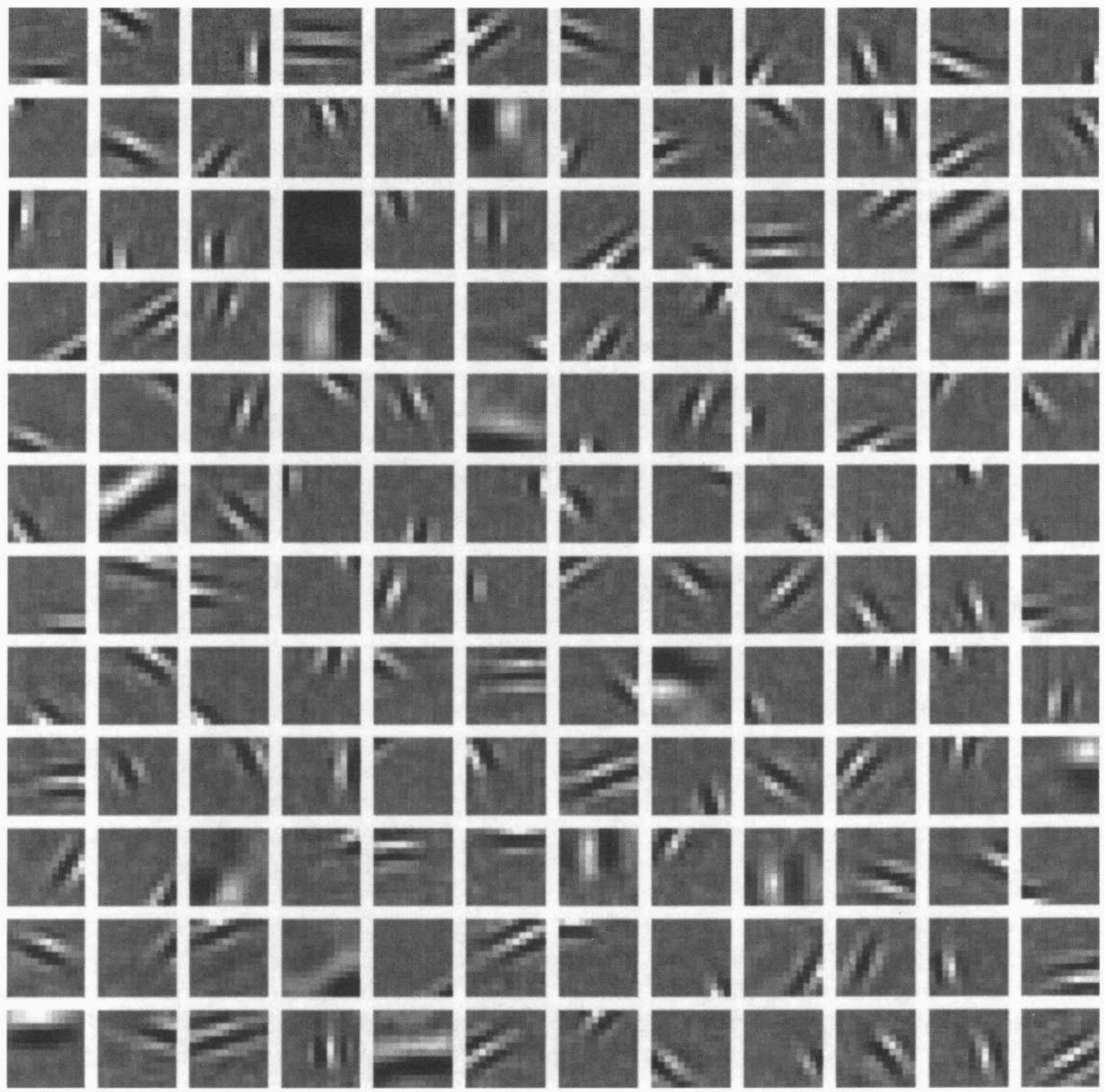
Understanding vision through hand-crafted features



Standard model:

$$r(s) = f(s;\theta)$$

e.g. combinations of Gabor filters:



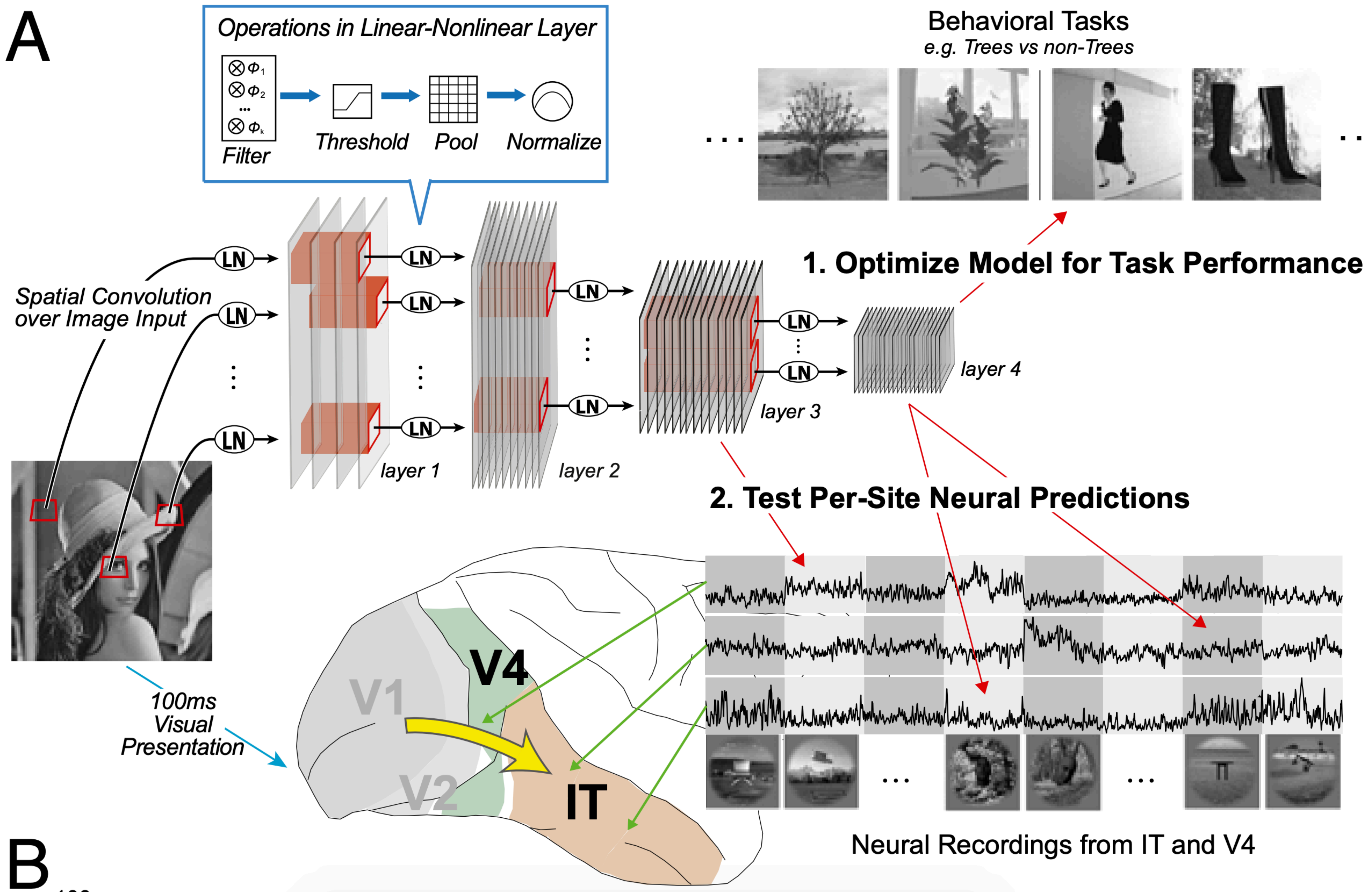
[Olshausen & Field, 1997]

ANN: let the model build itself

Performance-optimized hierarchical models predict neural responses in higher visual cortex

Daniel L. K. Yamins[a,1], Ha Hong[a,b,1], Charles F. Cadieu[a], Ethan A. Solomon[a], Darren Seibert[a], and James J. DiCarlo[a,2]

2014

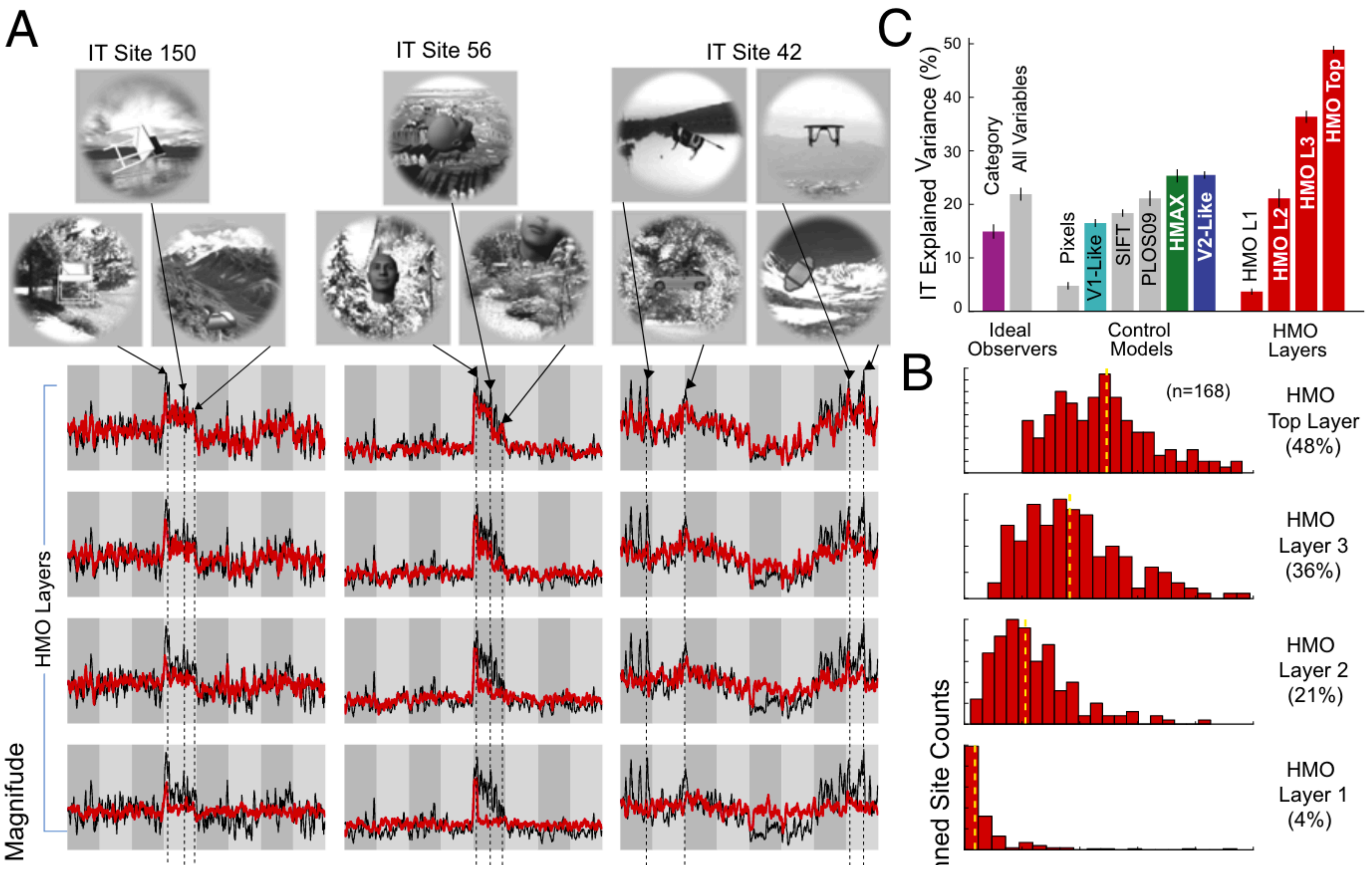


Model:

$$r(s) = f(\mathrm{net}(s); \theta)$$

Explaining IT cortex with ANNs

[Yamins et al 2014]



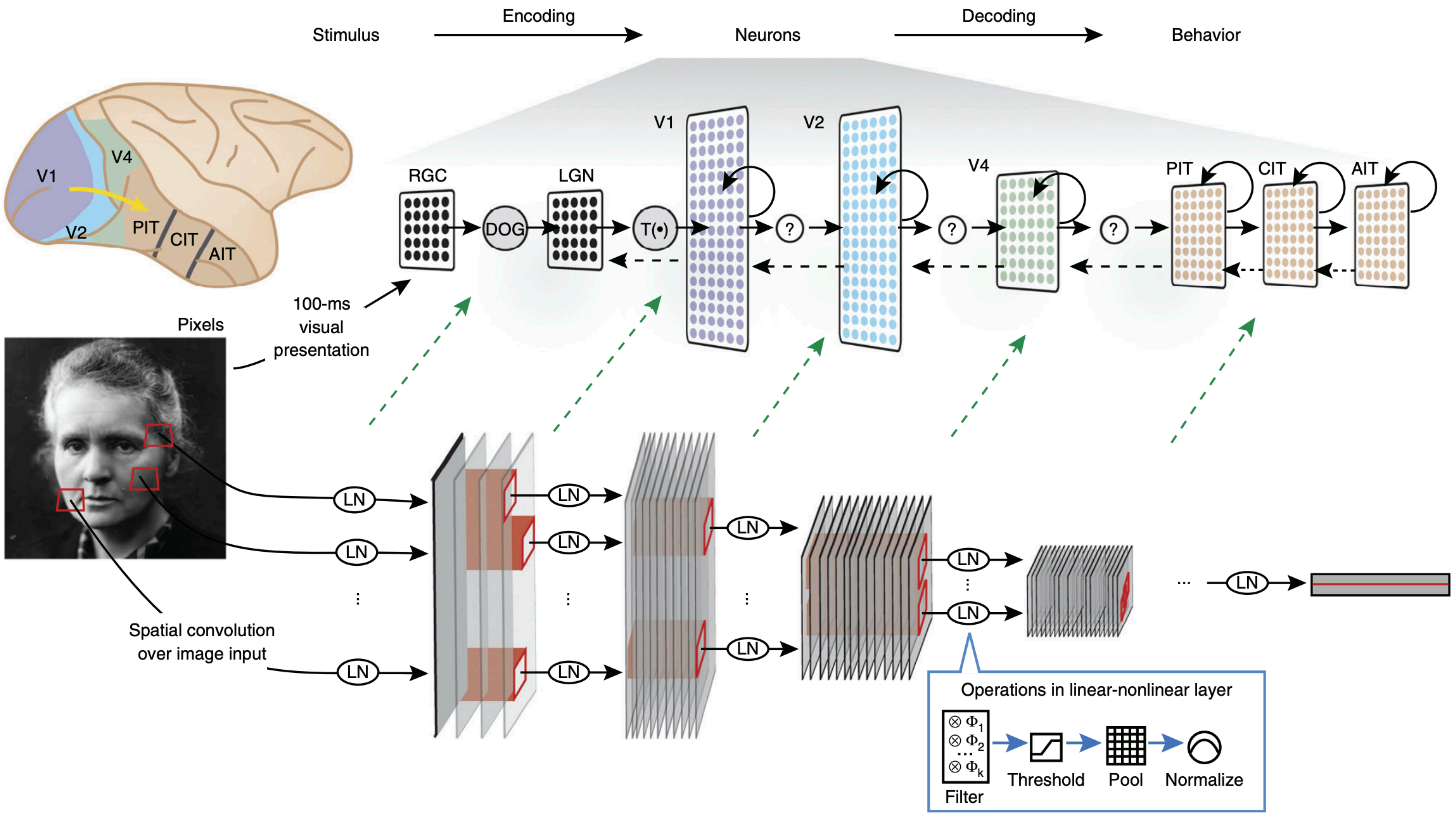
Explaining IT cortex with ANNs

Procedure:
1) train ANN on a behavioral task
2) compare ANN unit responses to real neural responses
3) compare that fit to alternative models (hand-crafted); to different network architectures / learning rules…

Note: the ANN was only trained behaviorally, not to match neural responses!

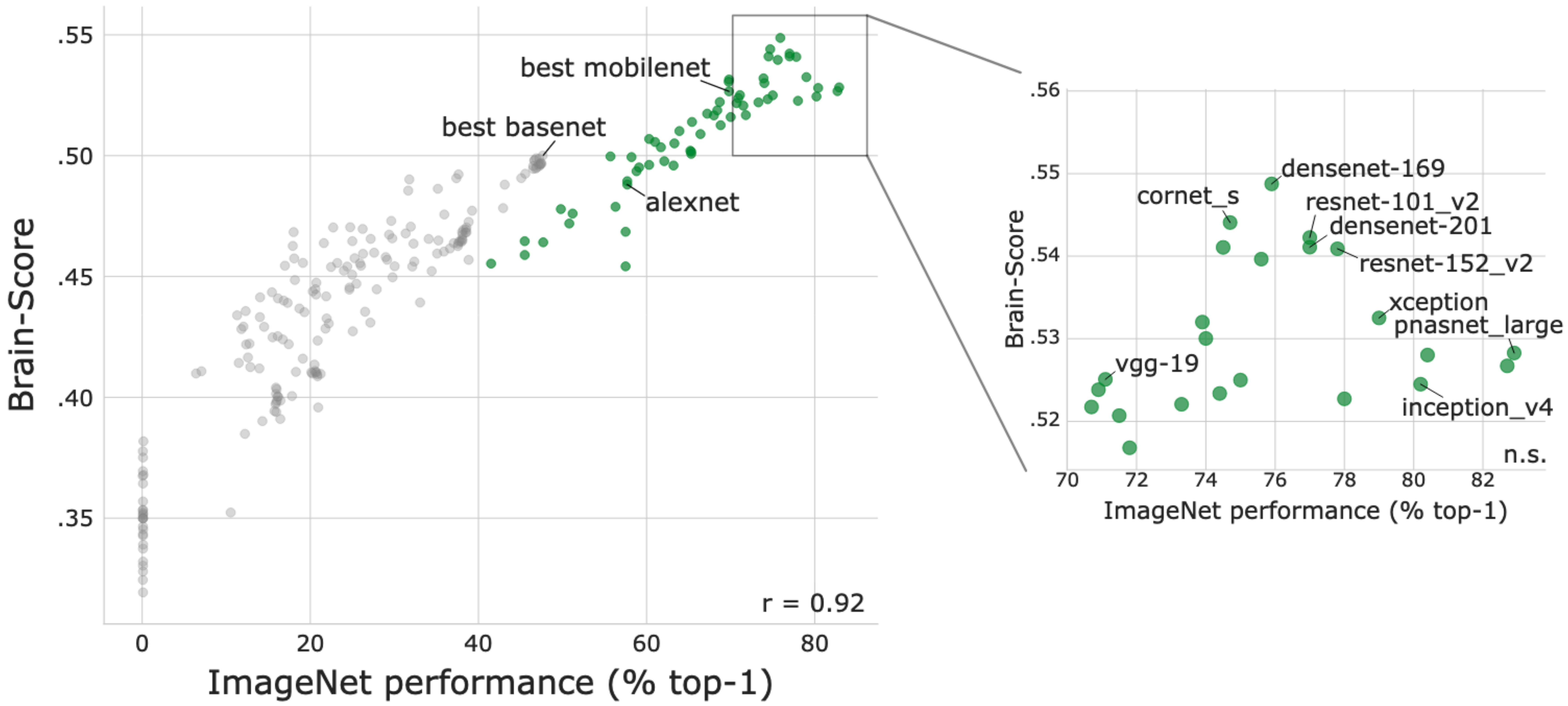
Insight 1: hierarchical learning in models and brains

[Yamins & DiCarlo 2016]



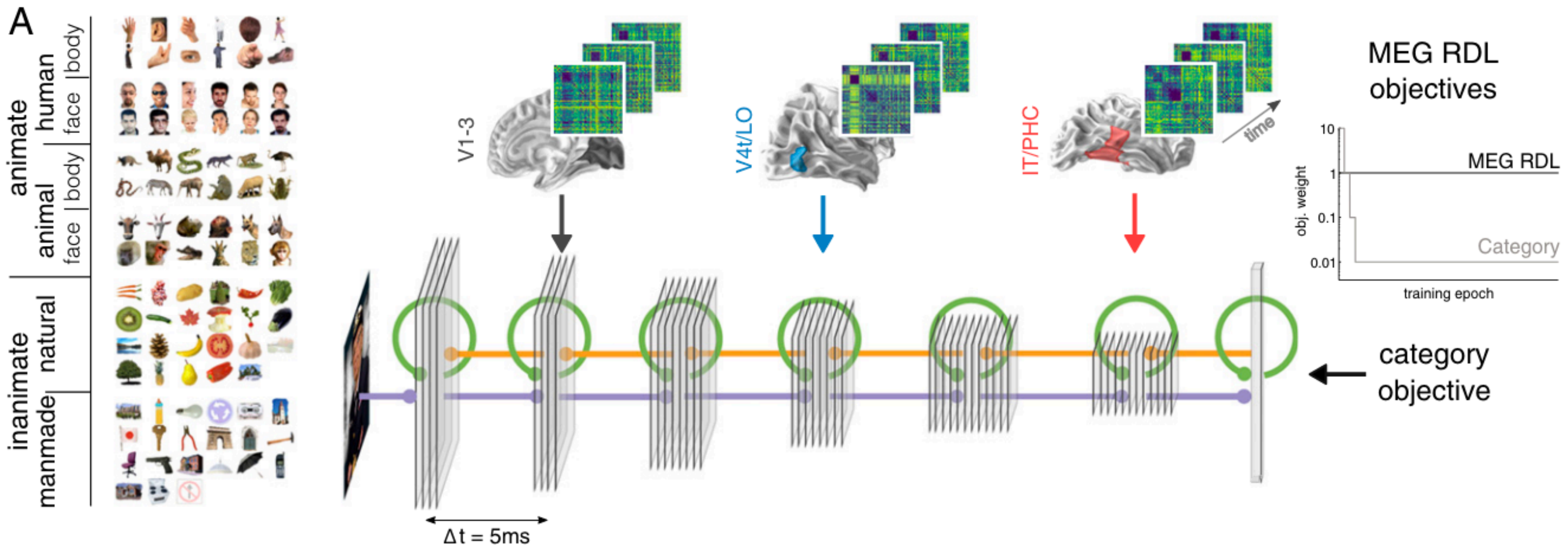
Insight 2: link between behavioral performance and neural fits

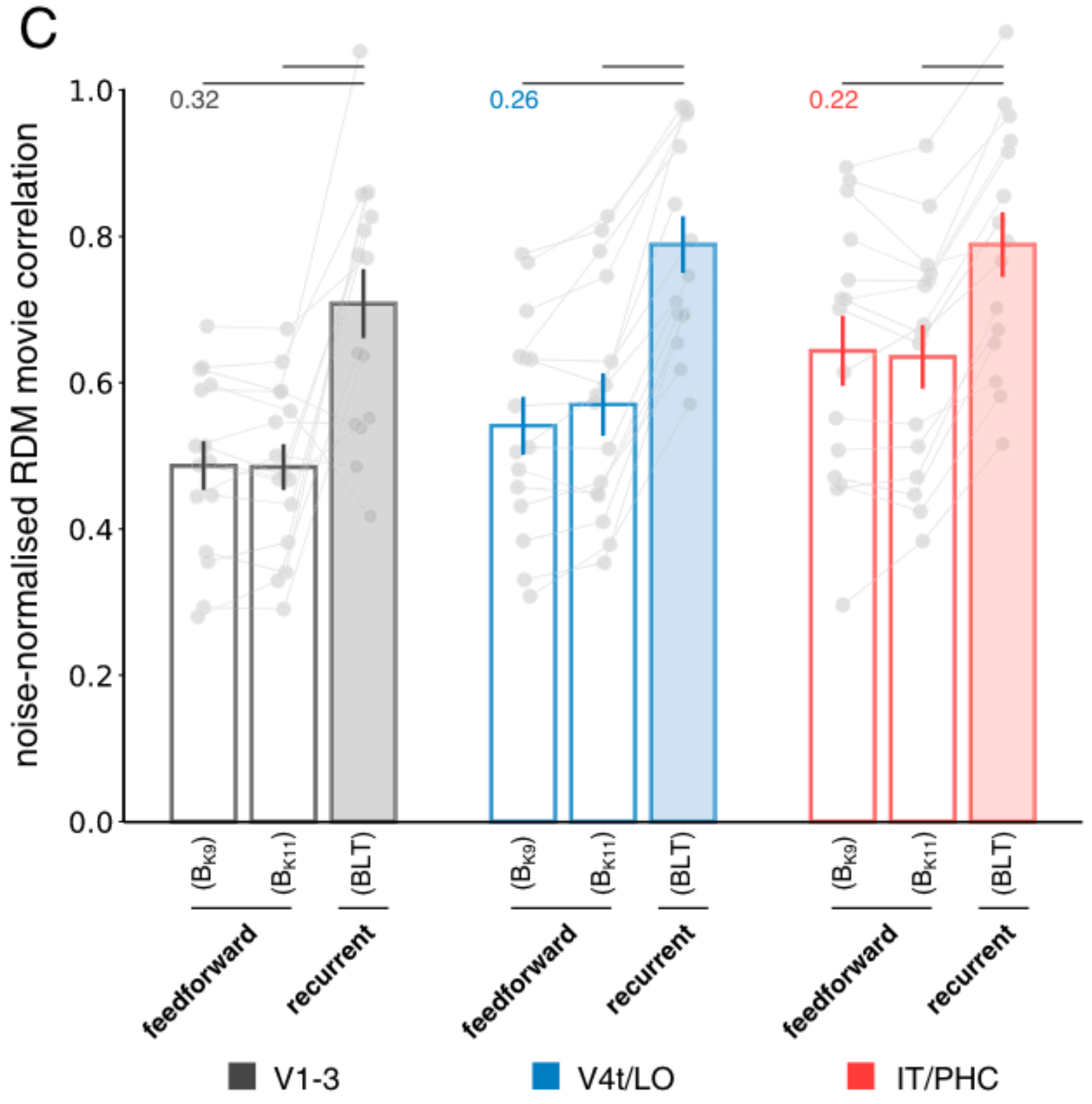
[Schrimpf et al 2020]



Insight 3: recurrence improves neural fits

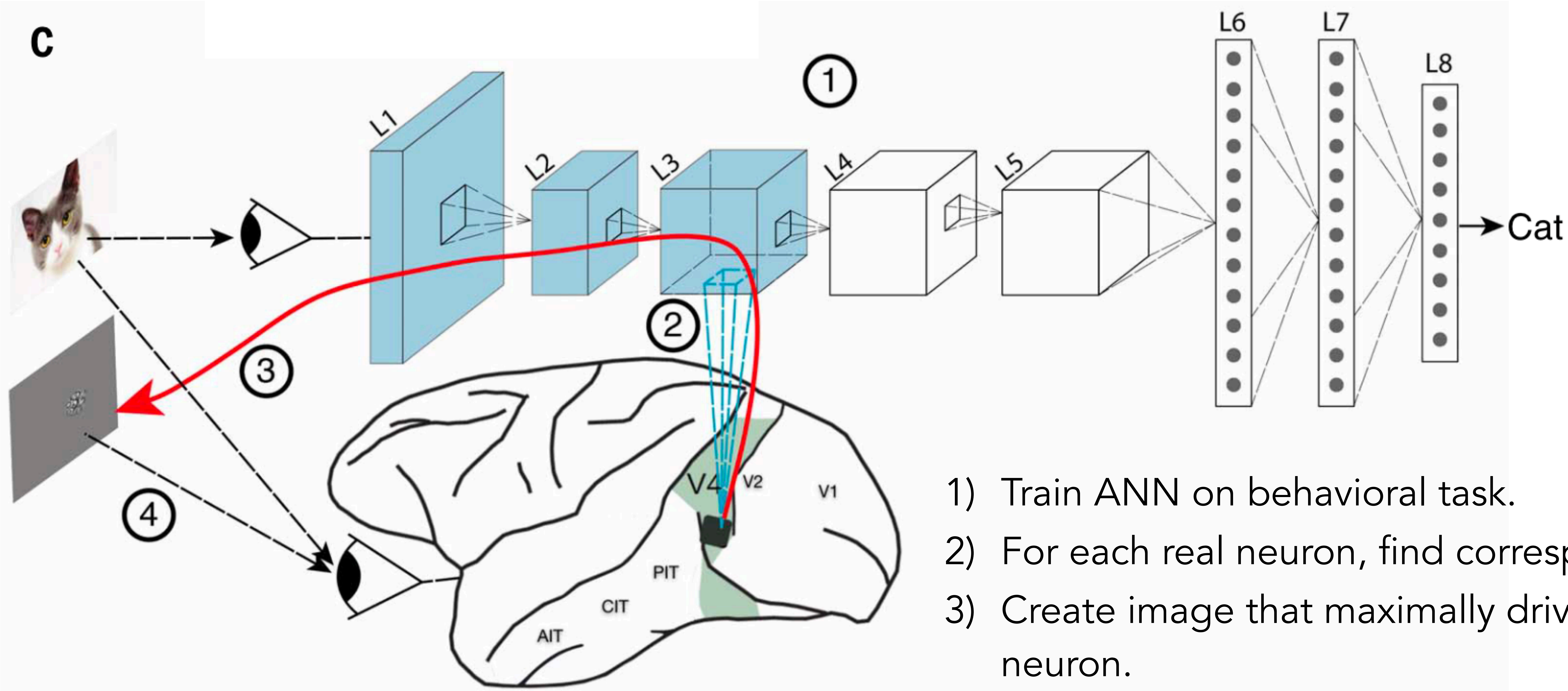
[Kietzmann et al., *Recurrence is required to capture the representational dynamics of the human visual system*, 2019]





Application: neural control

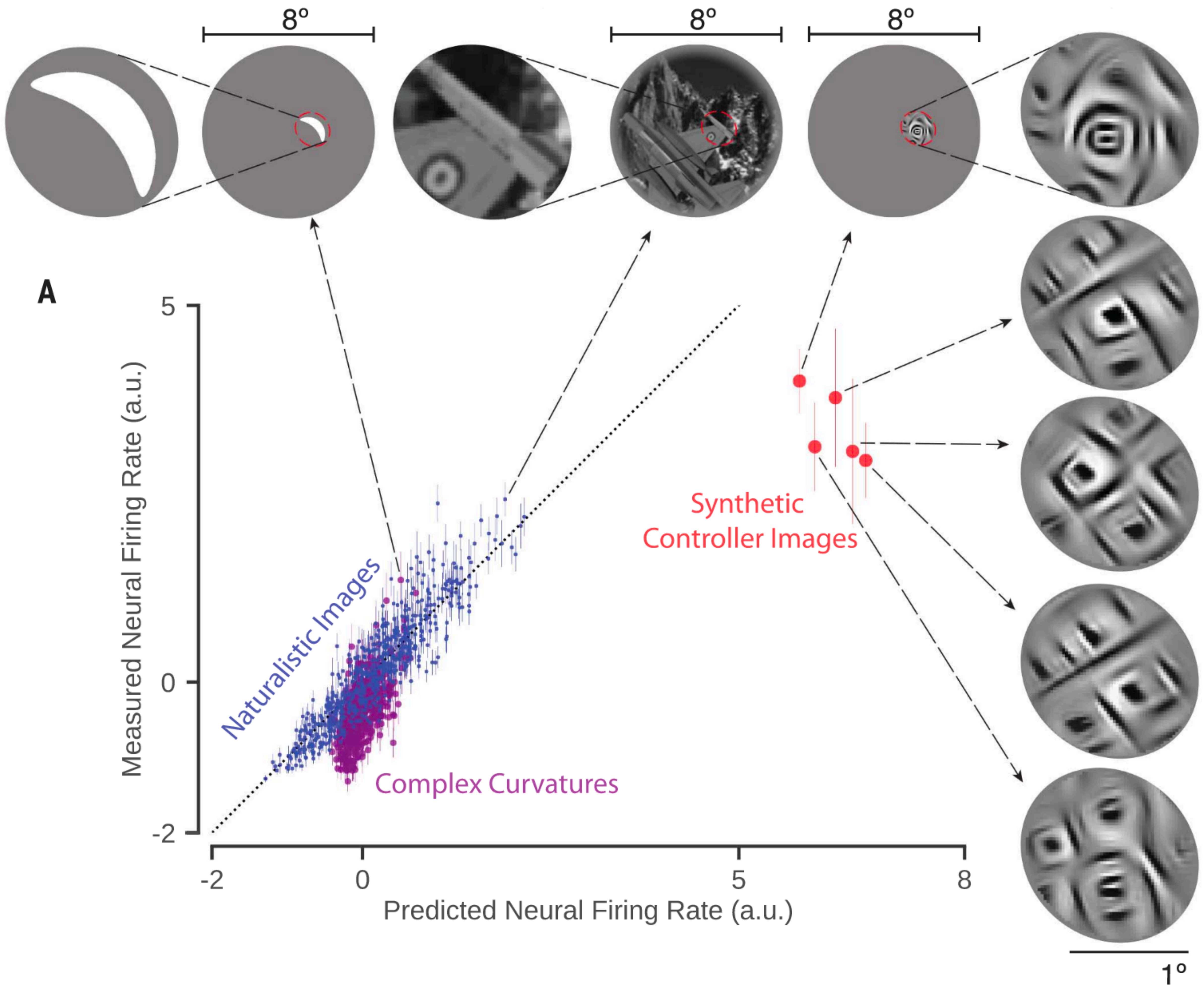
[Bashivan, Kar, DiCarlo, *Neural population control via deep image synthesis*, 2019]



1) Train ANN on behavioral task.
2) For each real neuron, find corresp. artificial neuron.
3) Create image that maximally drives the artificial neuron.
4) Check effect of that image on real neuron.

Application: neural control

[Bashivan, Kar, DiCarlo, *Neural population control via deep image synthesis*, 2019]



Critiques

- We replace a big, complicated thing by another big, complicated thing

    ➡ Yes, but the second one is *in silico*, and might be easier to understand.

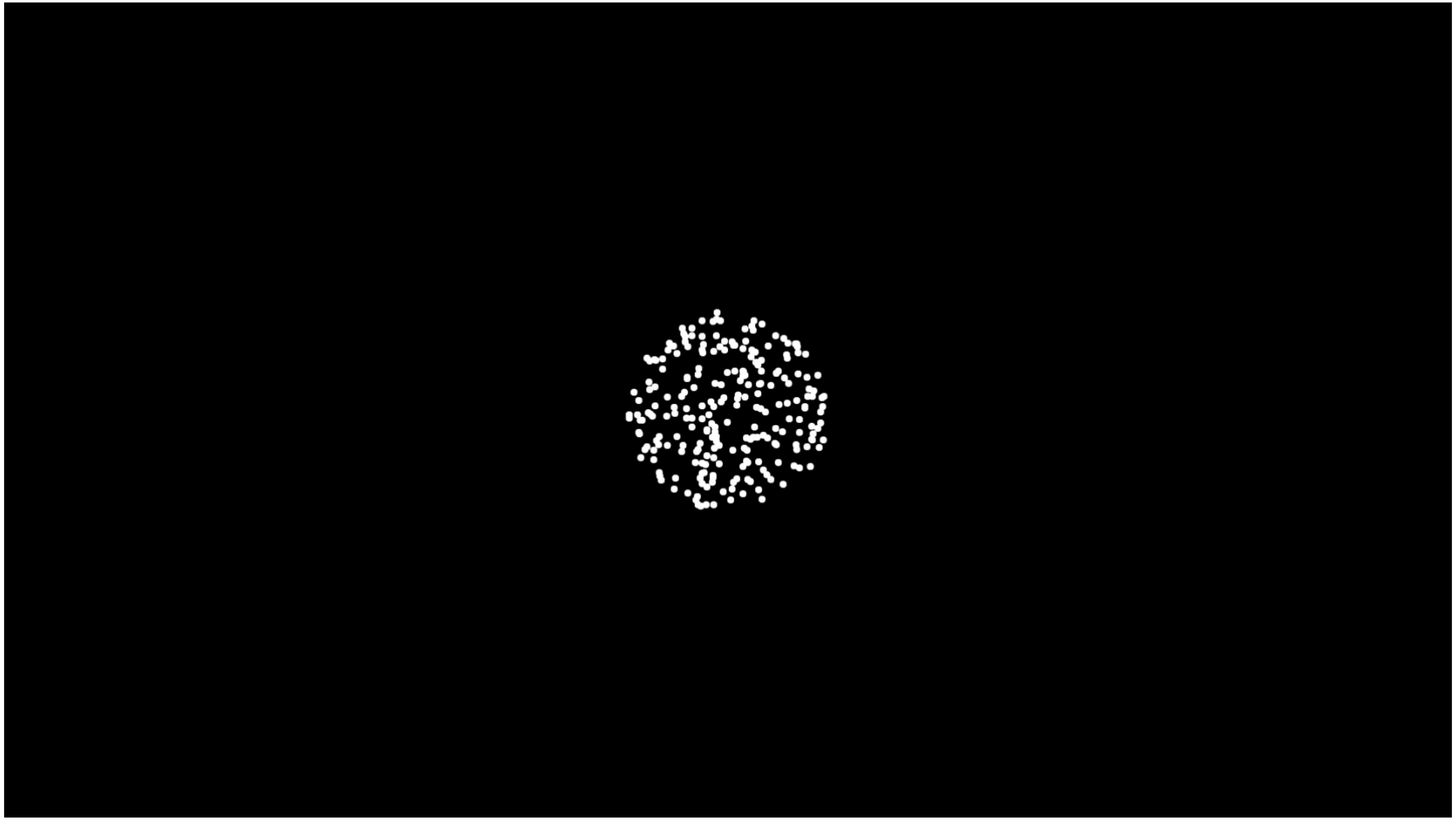
- The brain does not learn by backpropagation.

    ➡ Yes, but still relevant to understand neural function post-learning
    ➡ Plus, different training algorithms can be tried on ANNs and compared with brain data.

[A. Saxe, *If deep learning is the answer, then what is the question ?*, 2021]

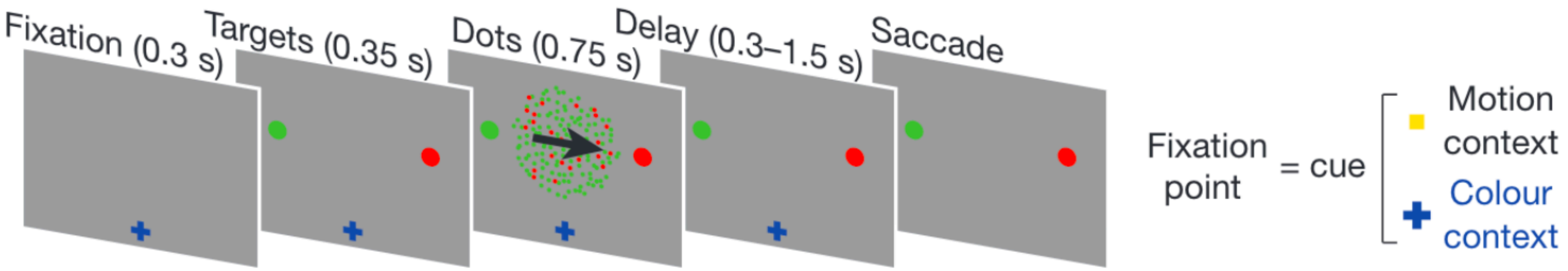
II - Temporal computations

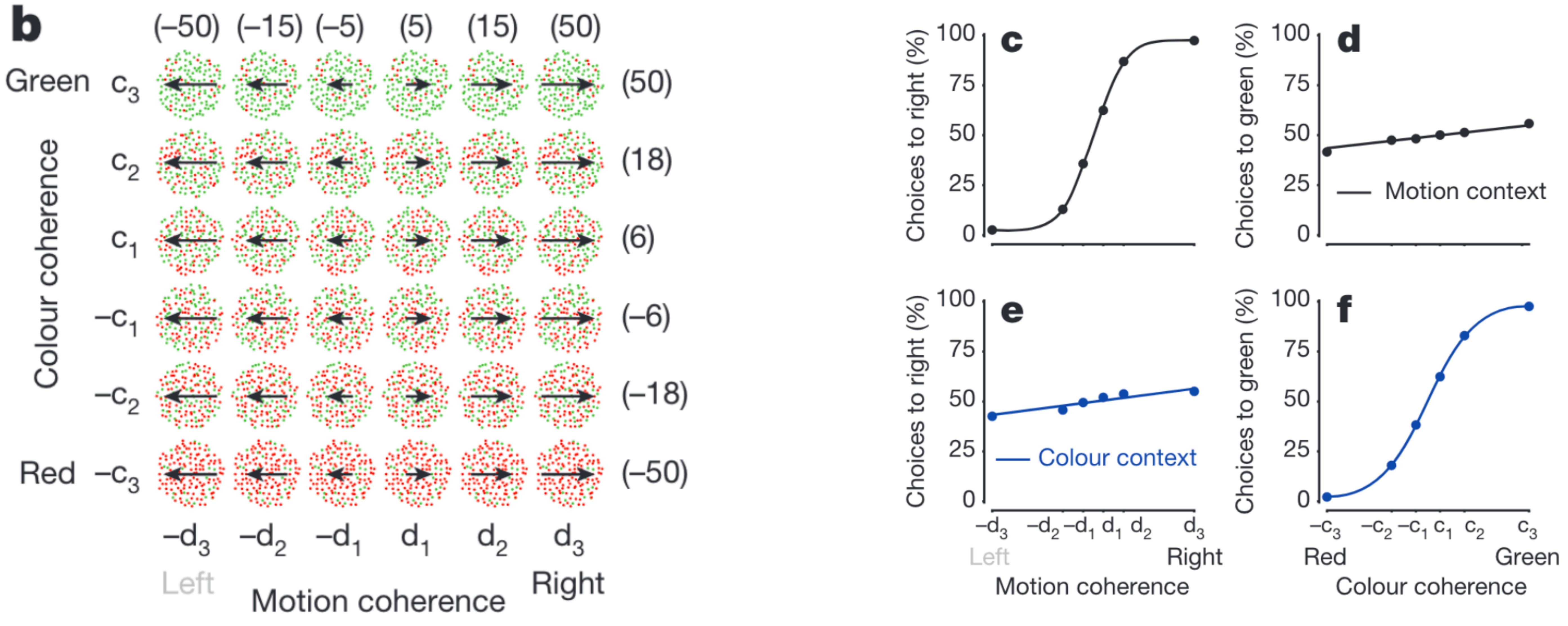
Decision-making tasks



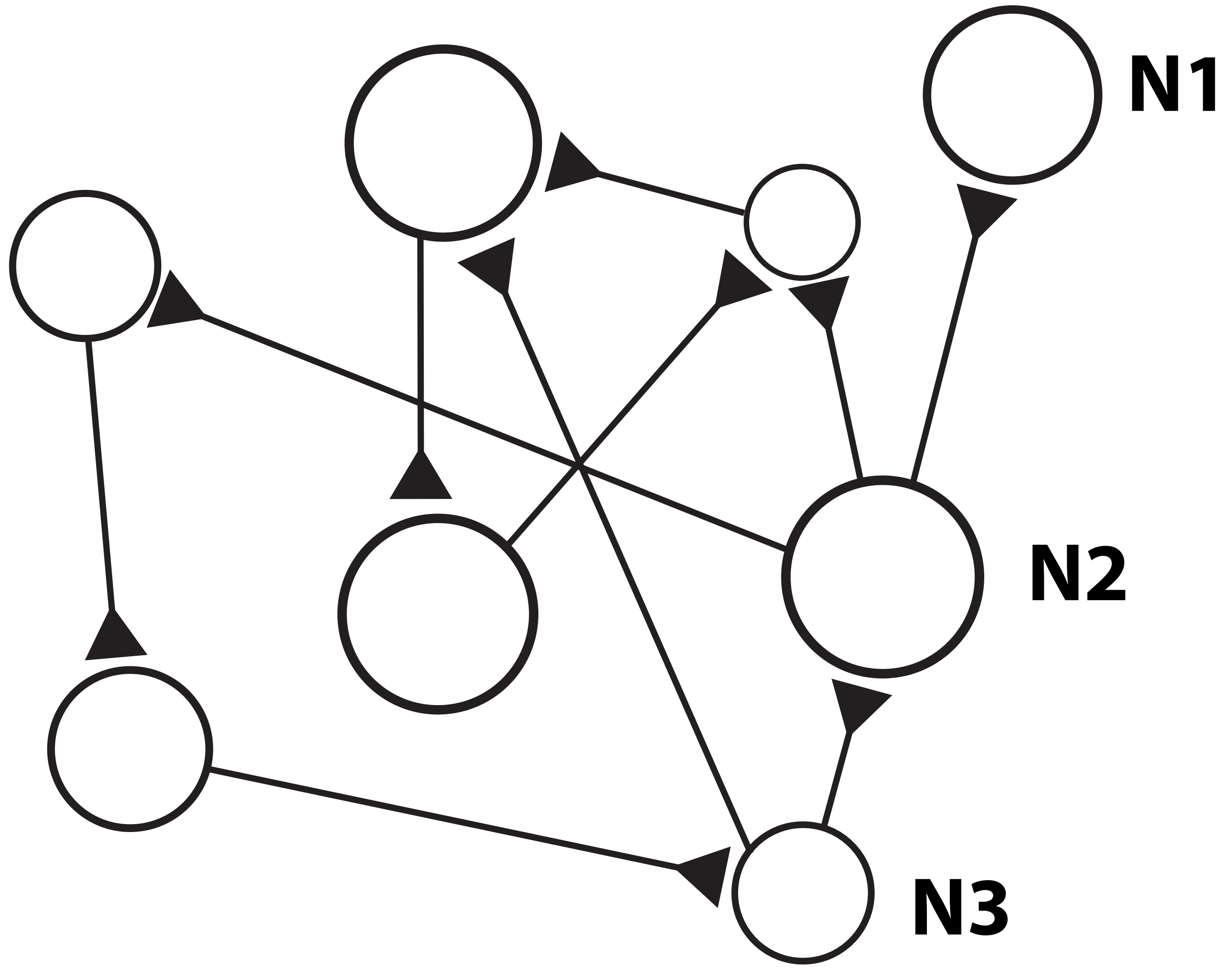
Cognition and RNNs

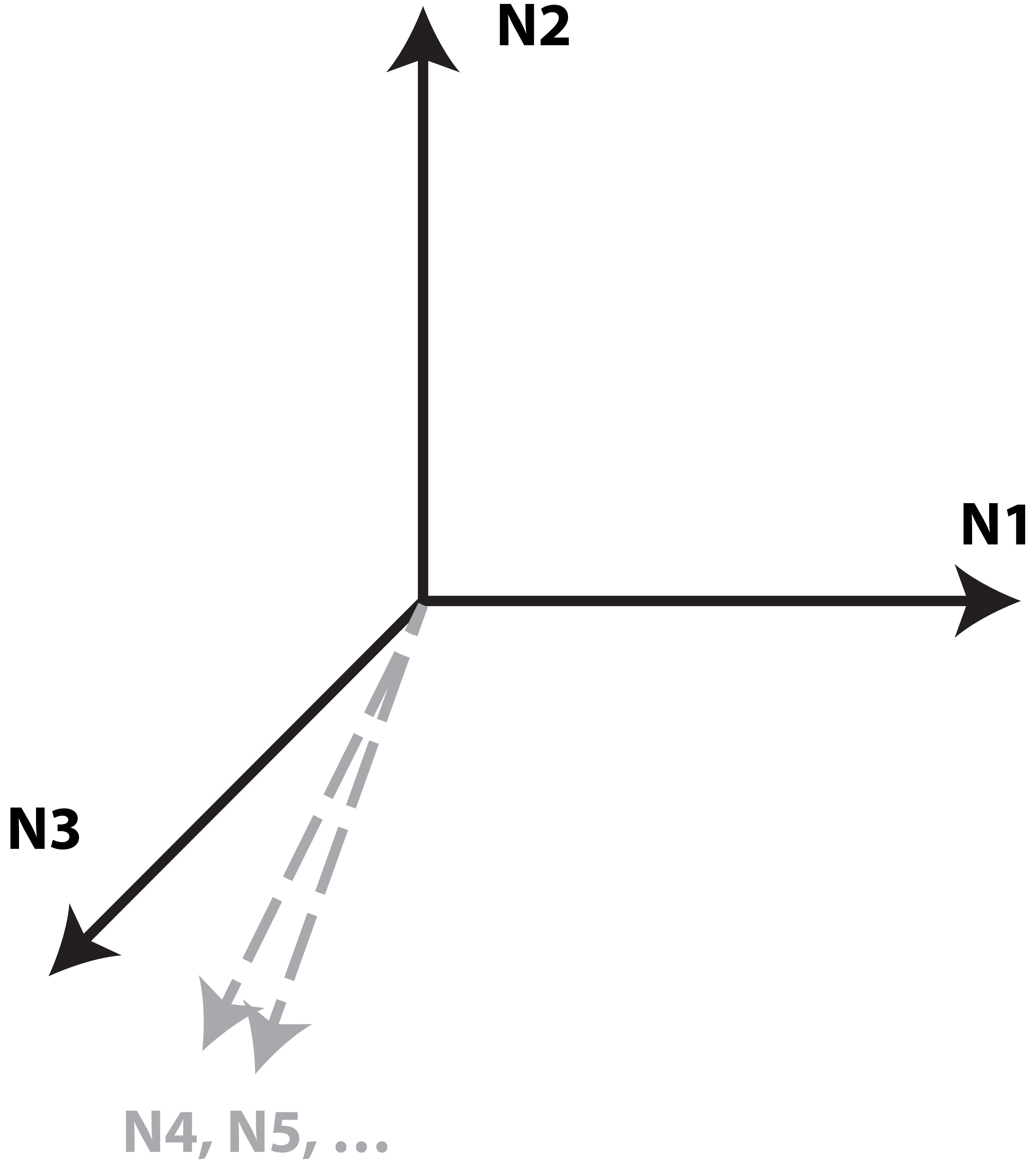
[Mante, Sussillo et al., *Context-dependent computations by recurrent dynamics in prefrontal cortex*, 2013]



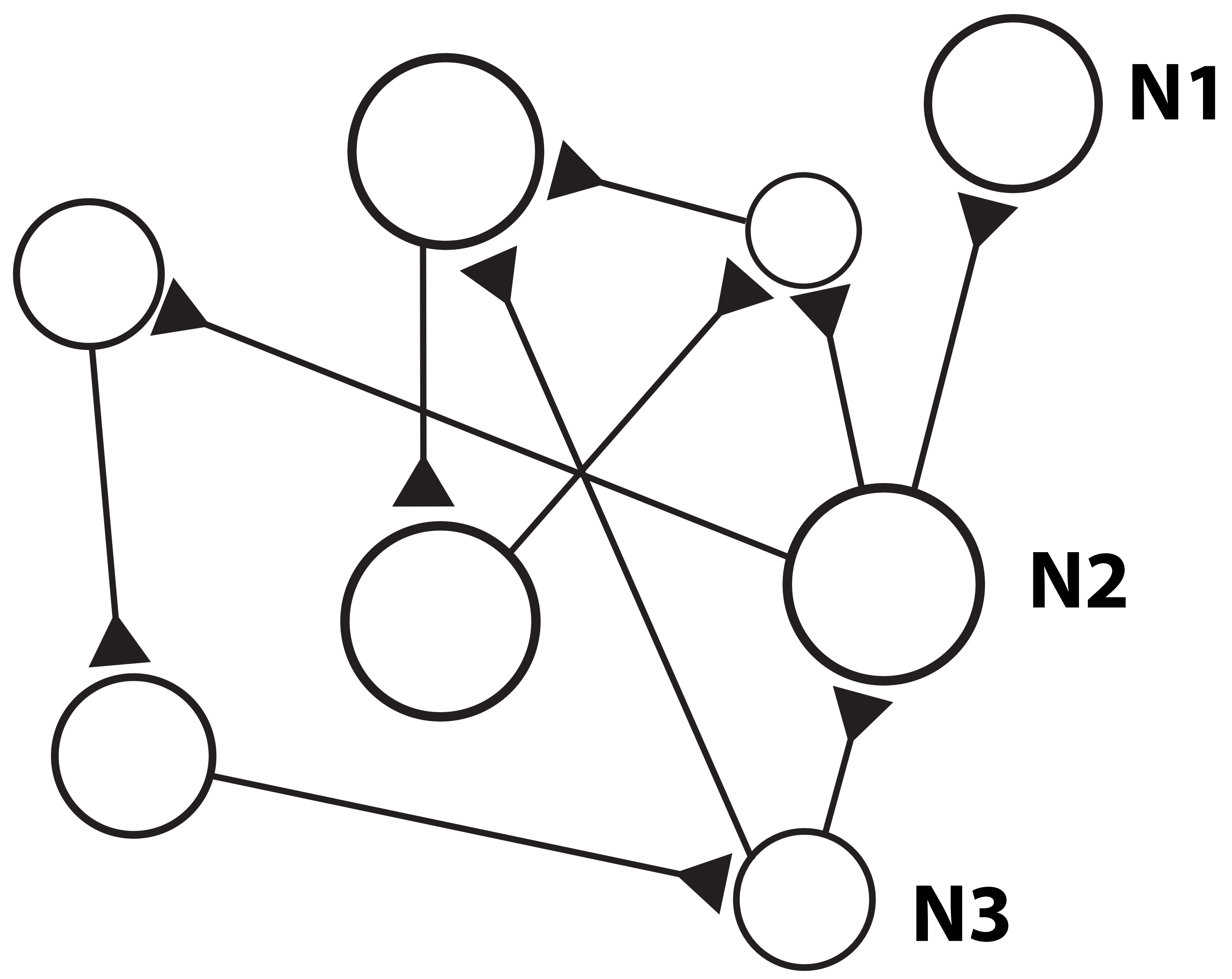


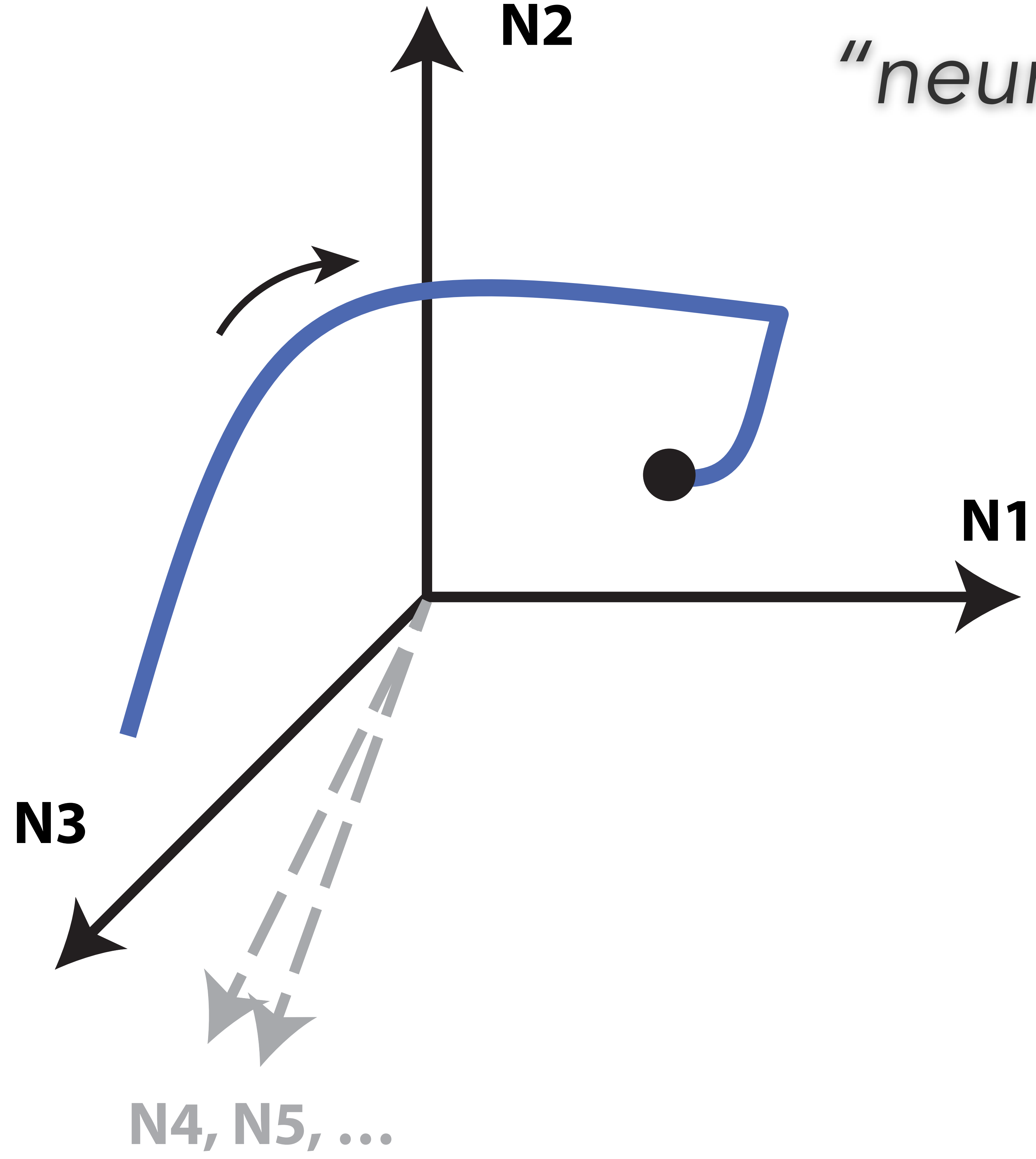
State-space analysis





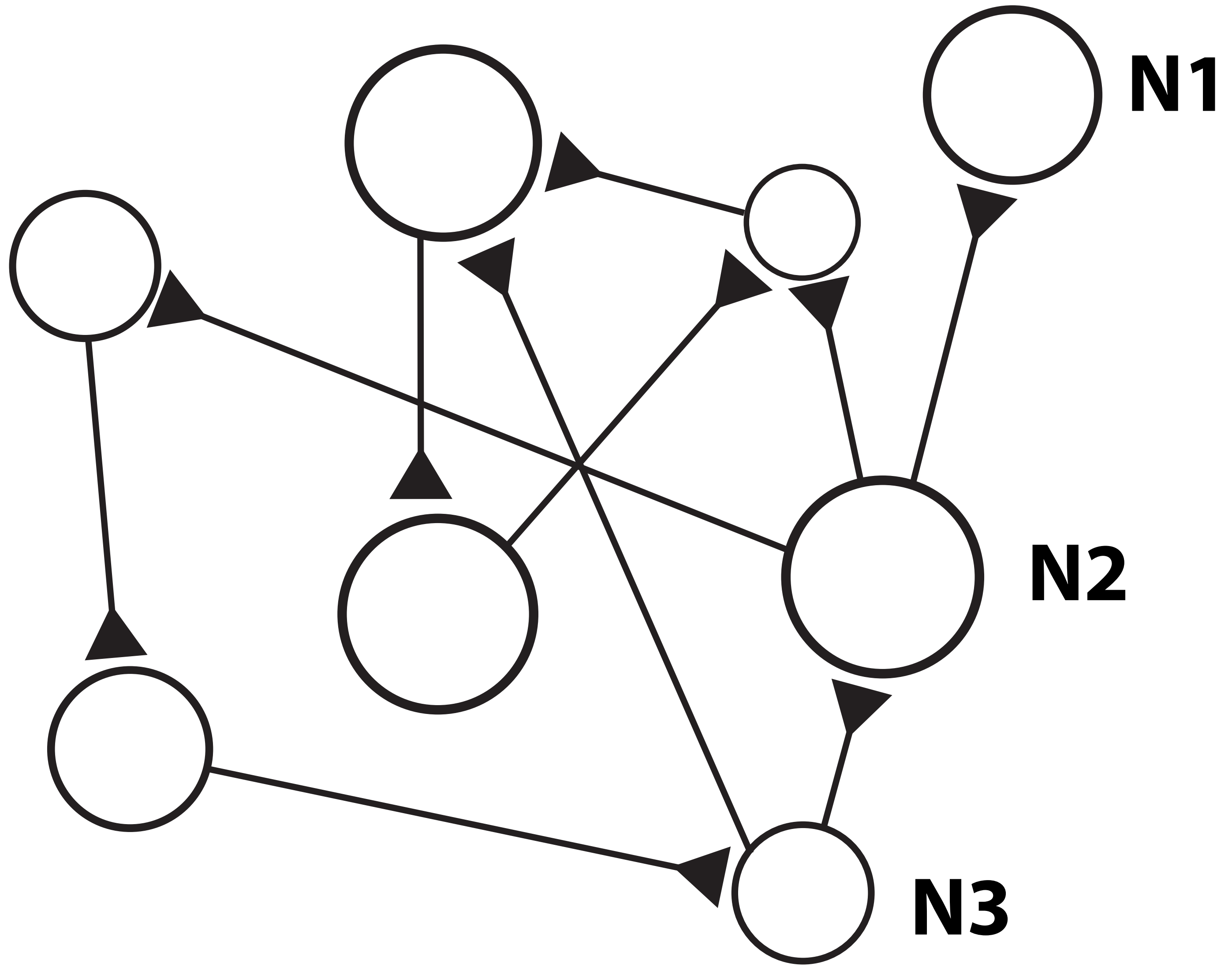
State-space analysis

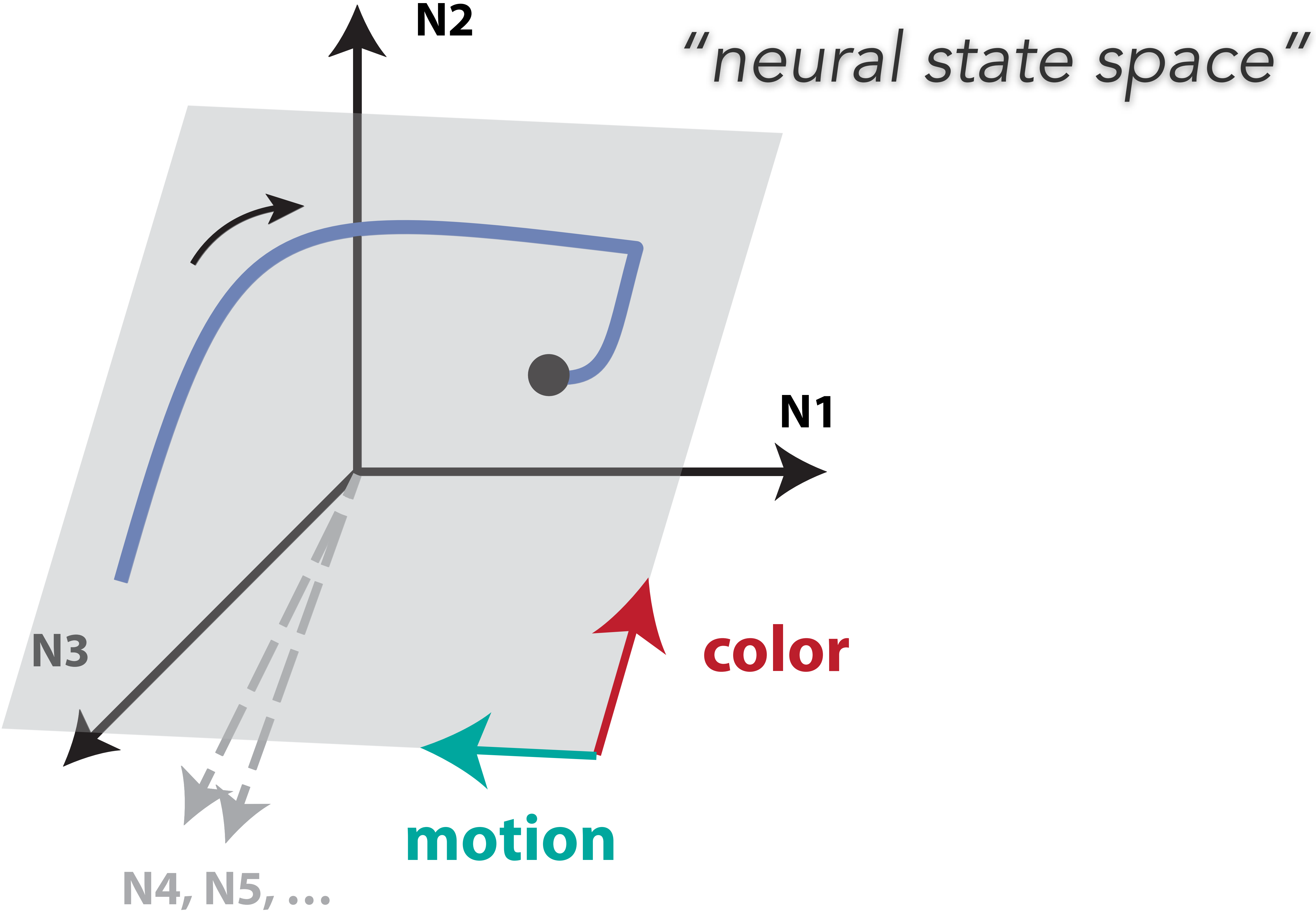




"neural state space"

State space analysis

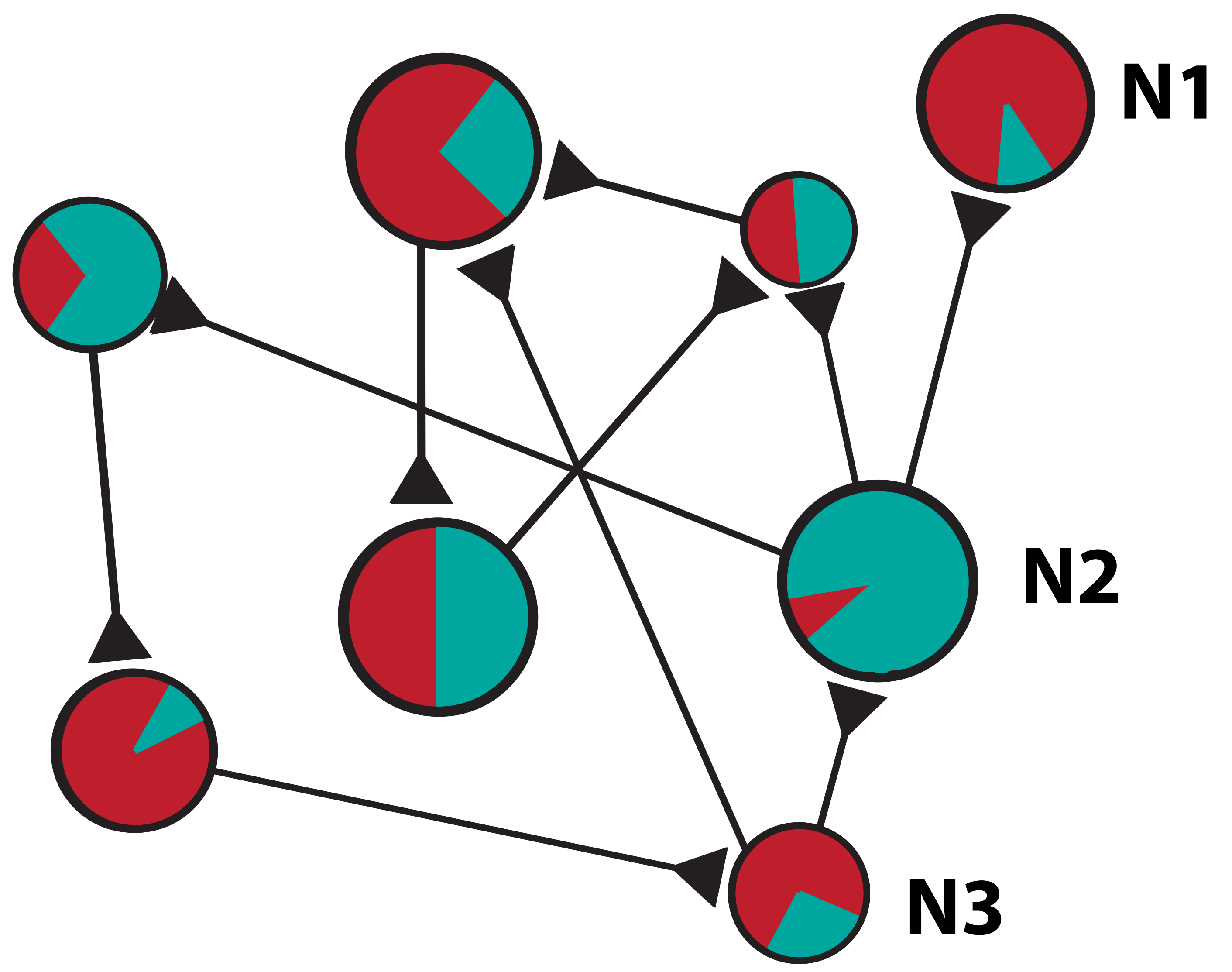


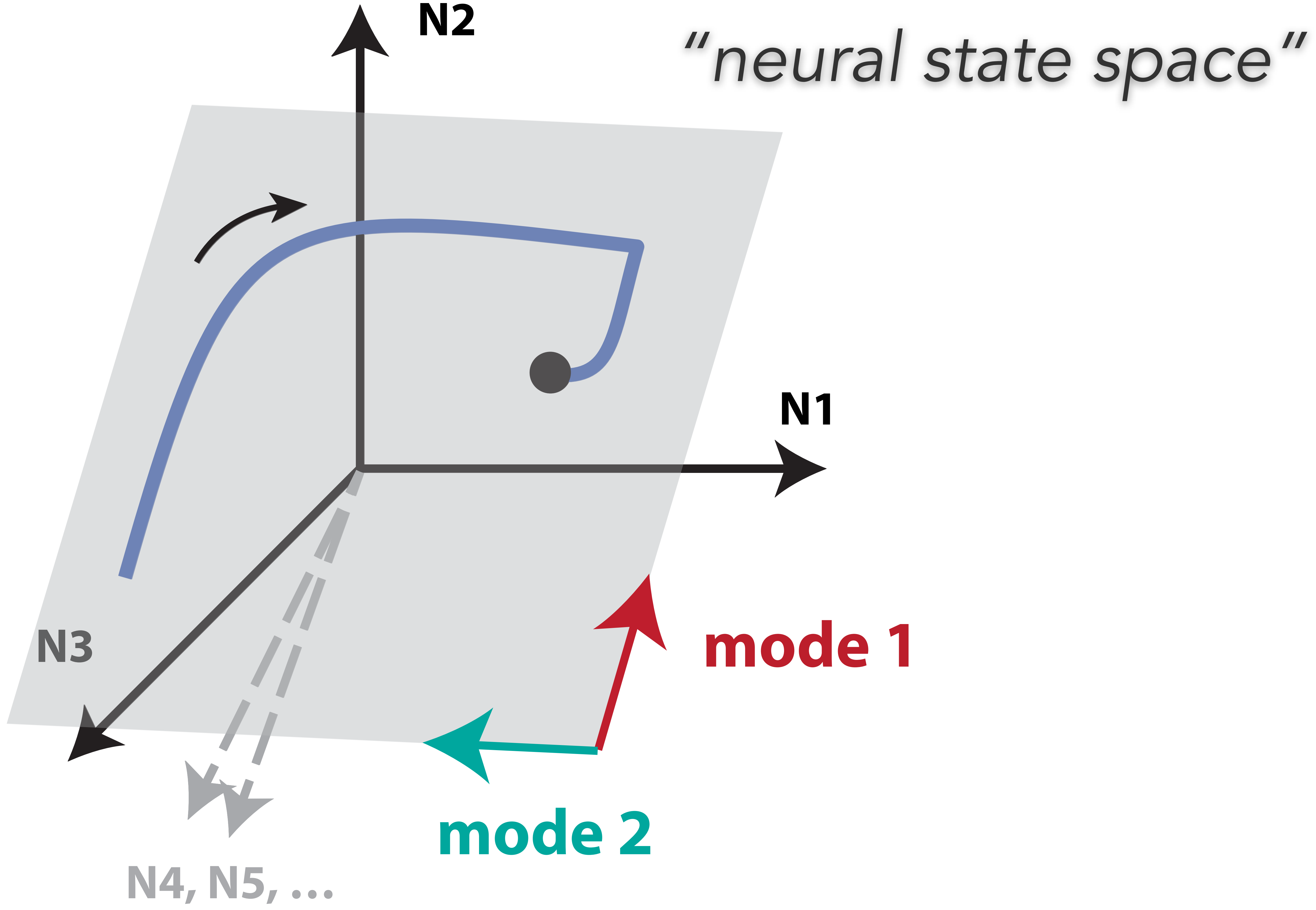


State space analysis

[Yuste, Nat Rev Neuro, 2015]

[Gallego et al., Neuron, 2017]



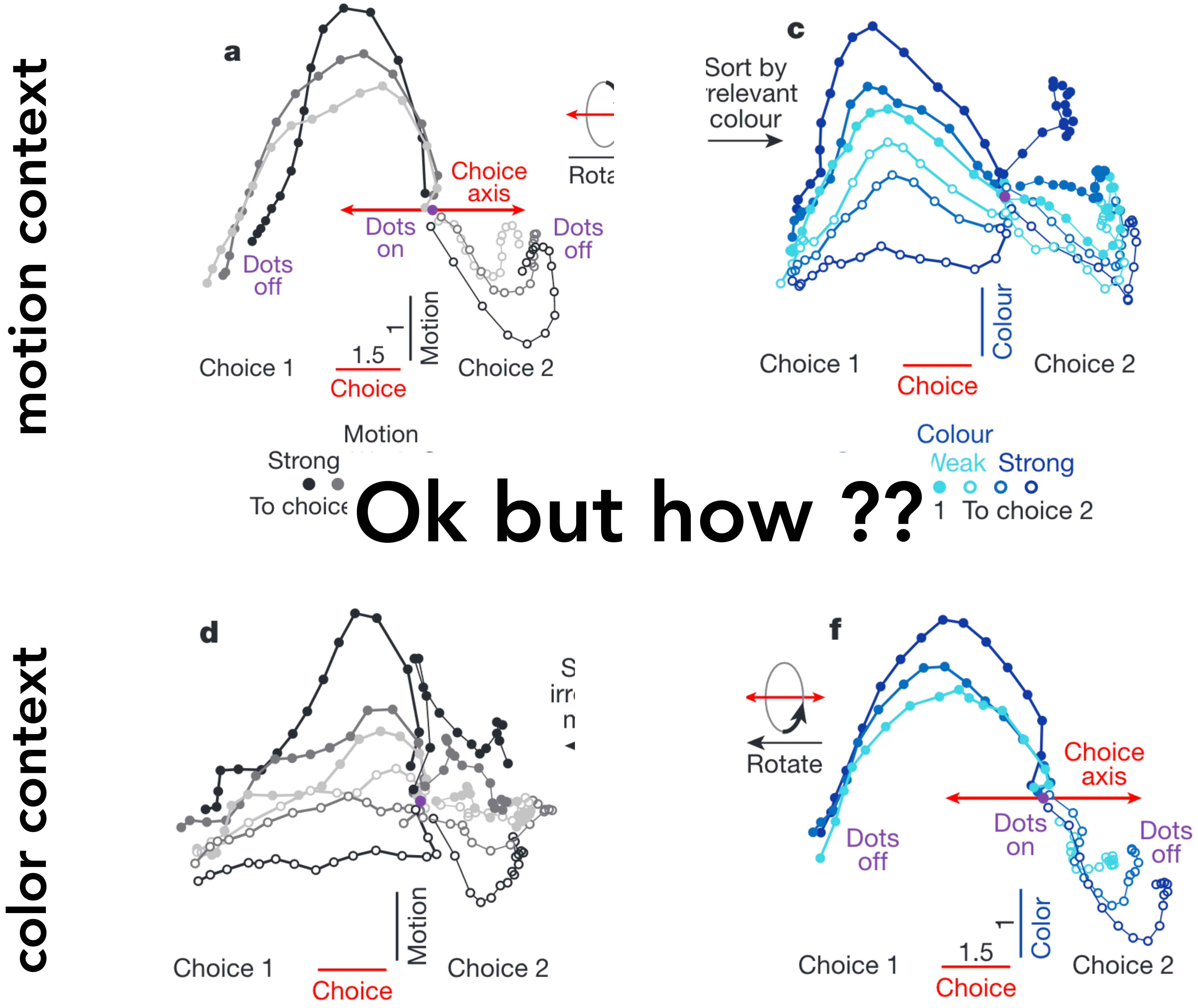


State space analysis: the Mante task

[Mante, Sussillo et al., *Context-dependent computations by recurrent dynamics in prefrontal cortex*, 2013]

Find 3 axes:

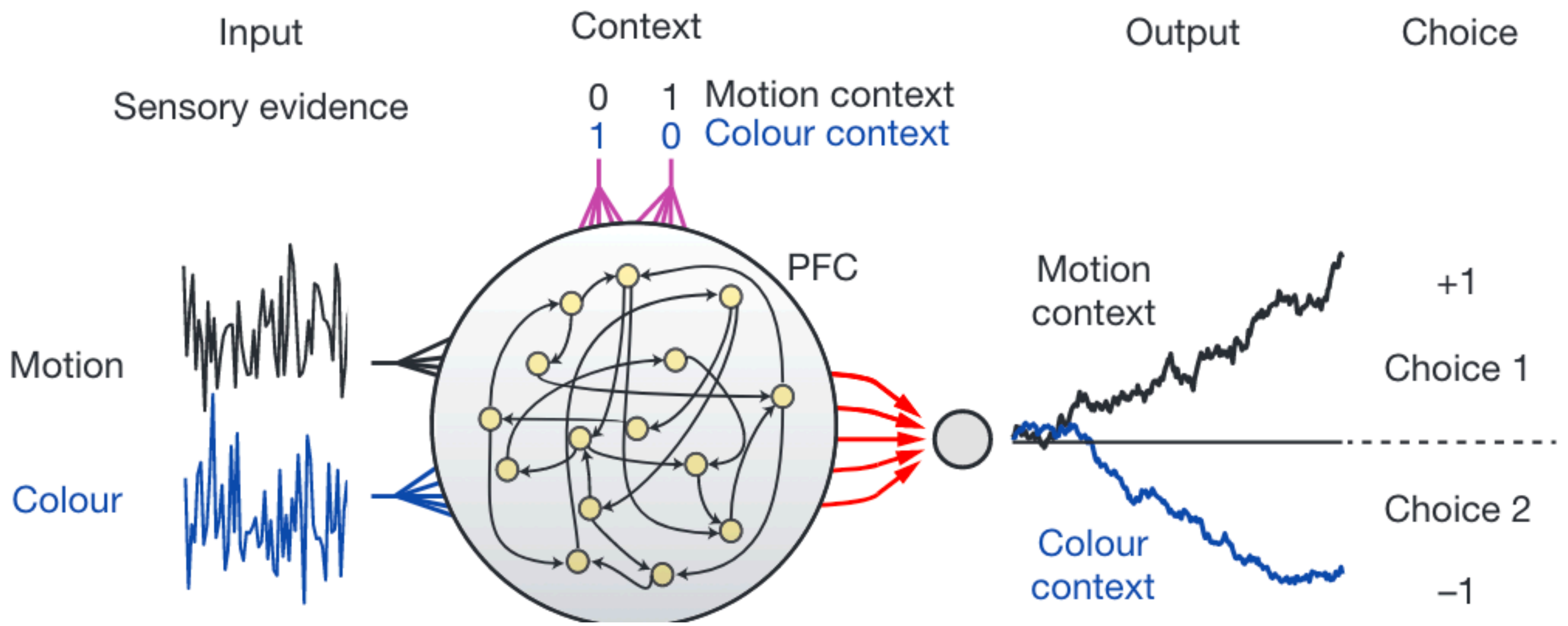
- color
- motion
- choice



Ok but how ??

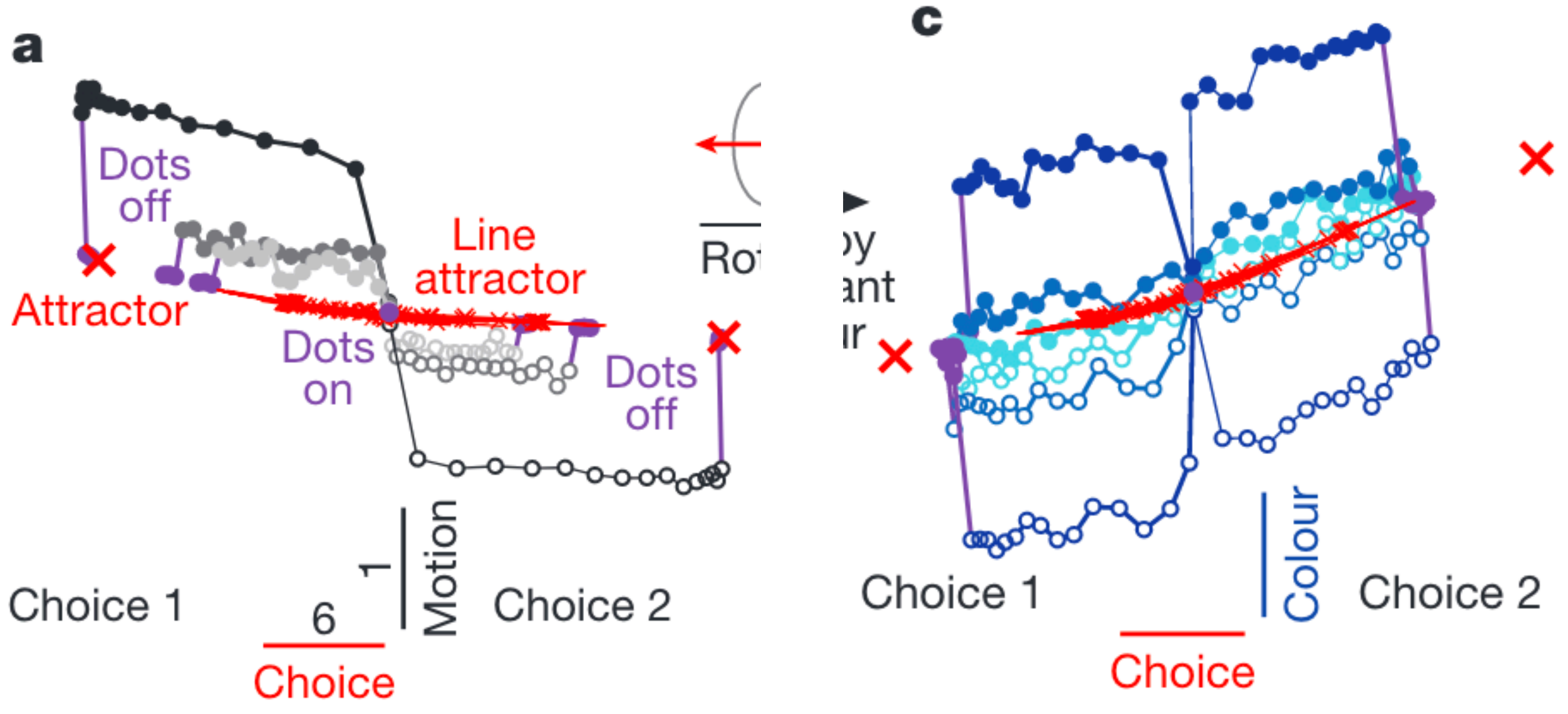
What would an RNN do ?

[Mante, Sussillo et al., *Context-dependent computations by recurrent dynamics in prefrontal cortex*, 2013]

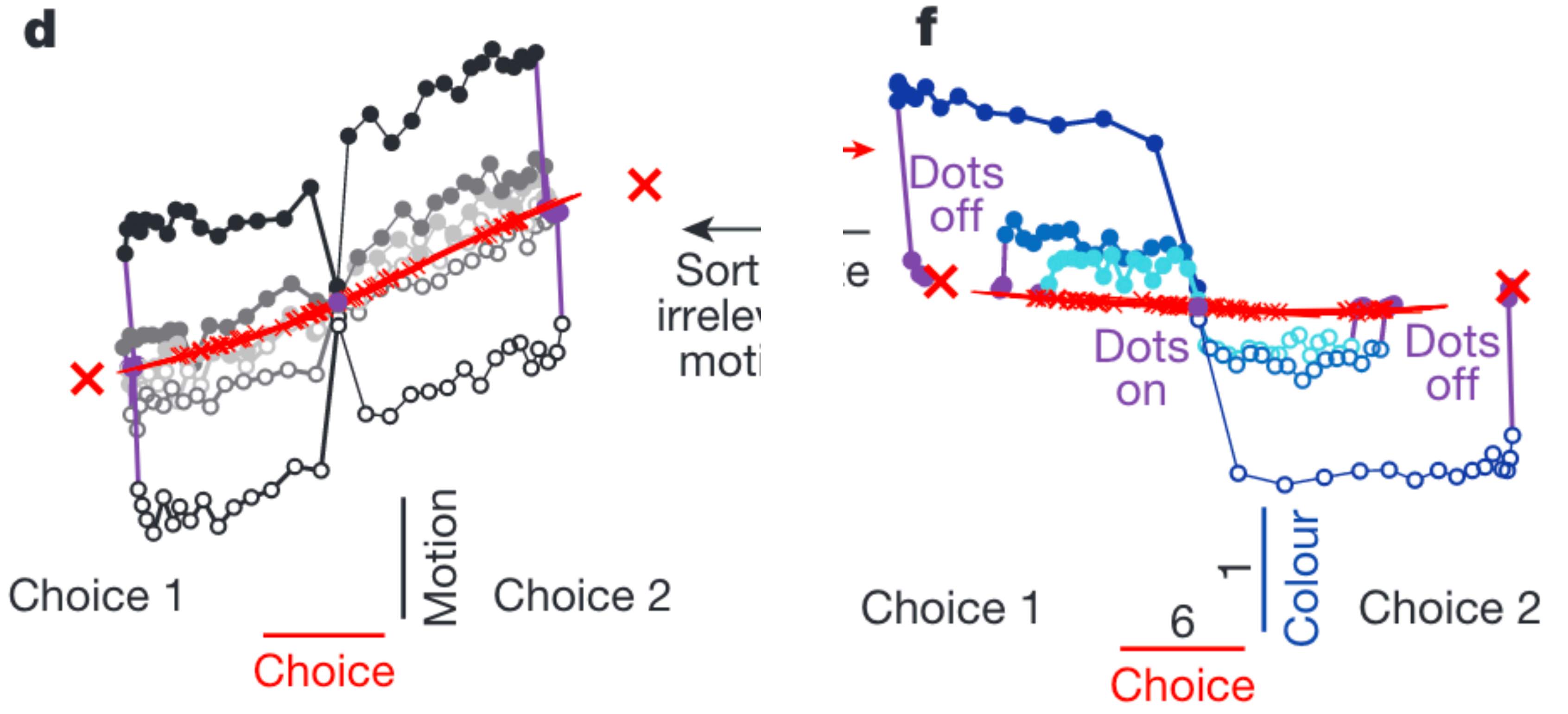


What would an RNN do ?

[Mante, Sussillo et al., *Context-dependent computations by recurrent dynamics in prefrontal cortex*, 2013]



➡ **Very similar dynamics as the monkey brain**



Notion of dynamical systems

Discrete **Continuous**

dx

Two formu

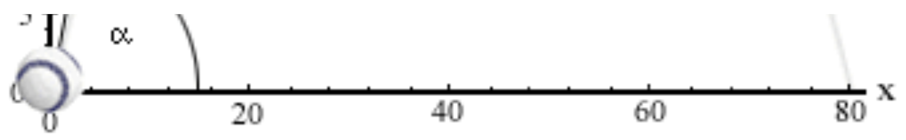
An RNN

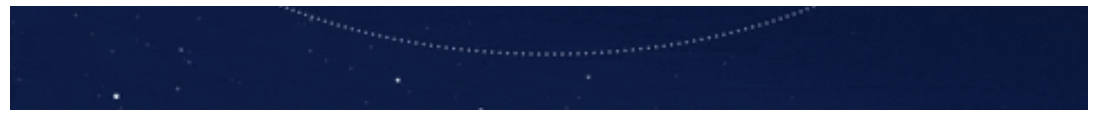
Other ex

- a ball

"Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective positions of the beings which compose it, if moreover this intelligence were vast enough to submit these data to analysis, it would embrace in the same formula both the movements of the largest bodies in the universe and those of the lightest atom; to it nothing would be uncertain, and the future as the past would be present to its eye."

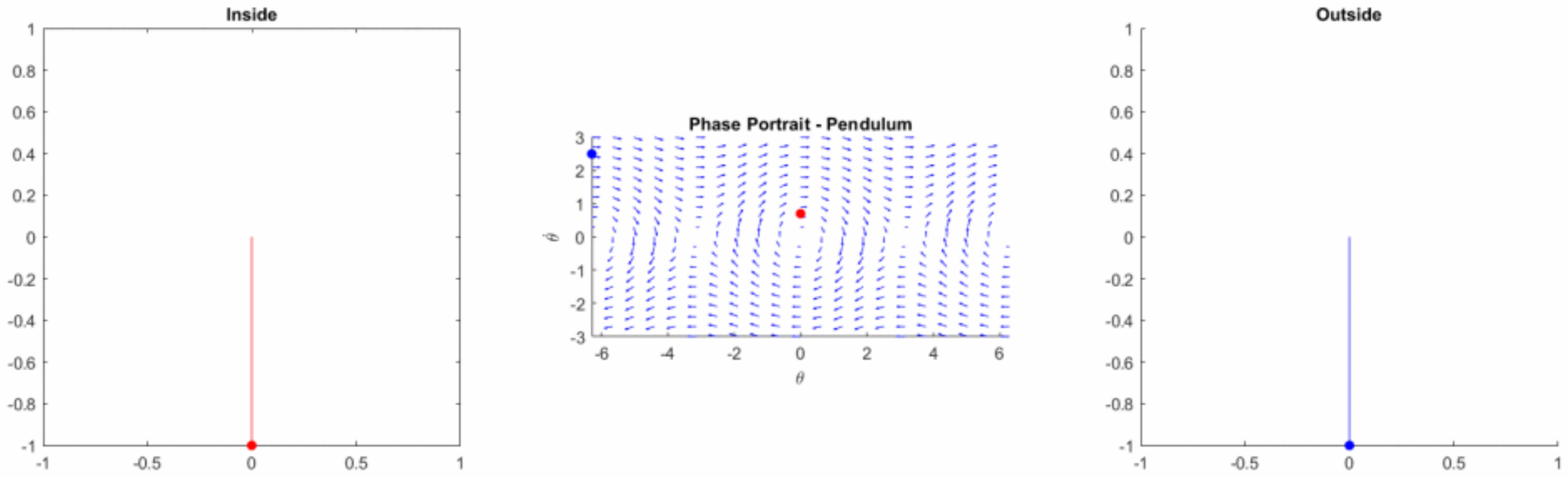
Pierre-Simon de Laplace





Flow fields

(aka phase portraits)

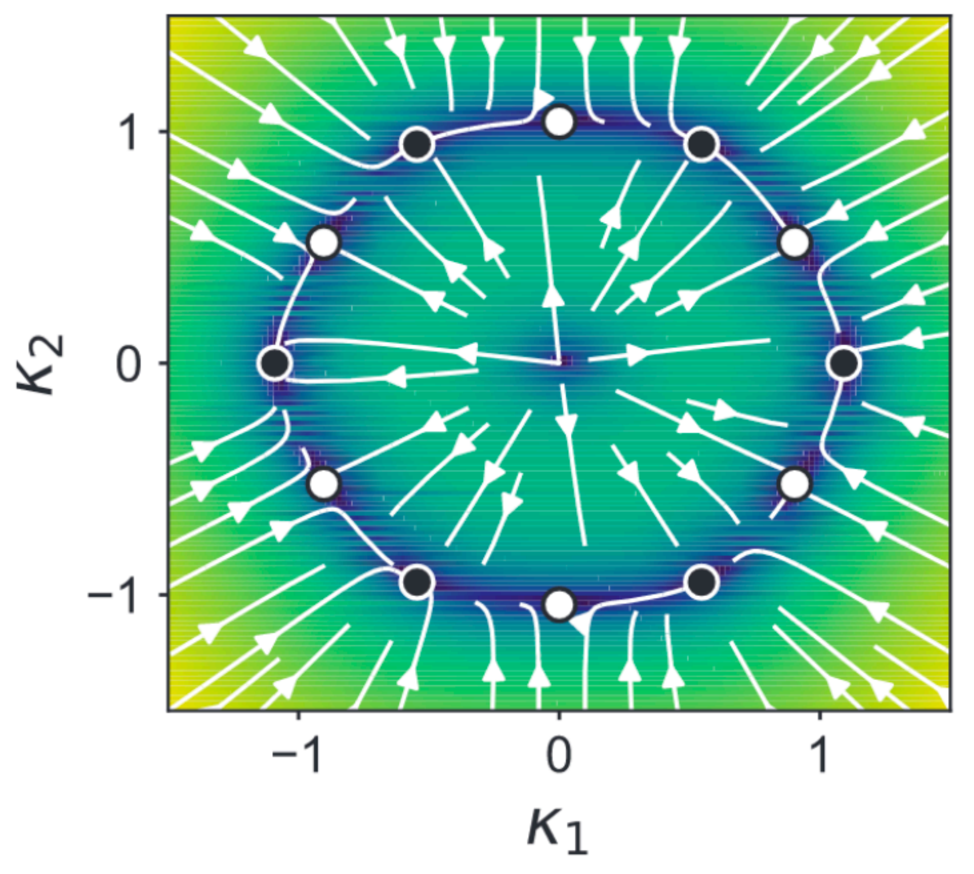


Long-term behavior of dynamical systems

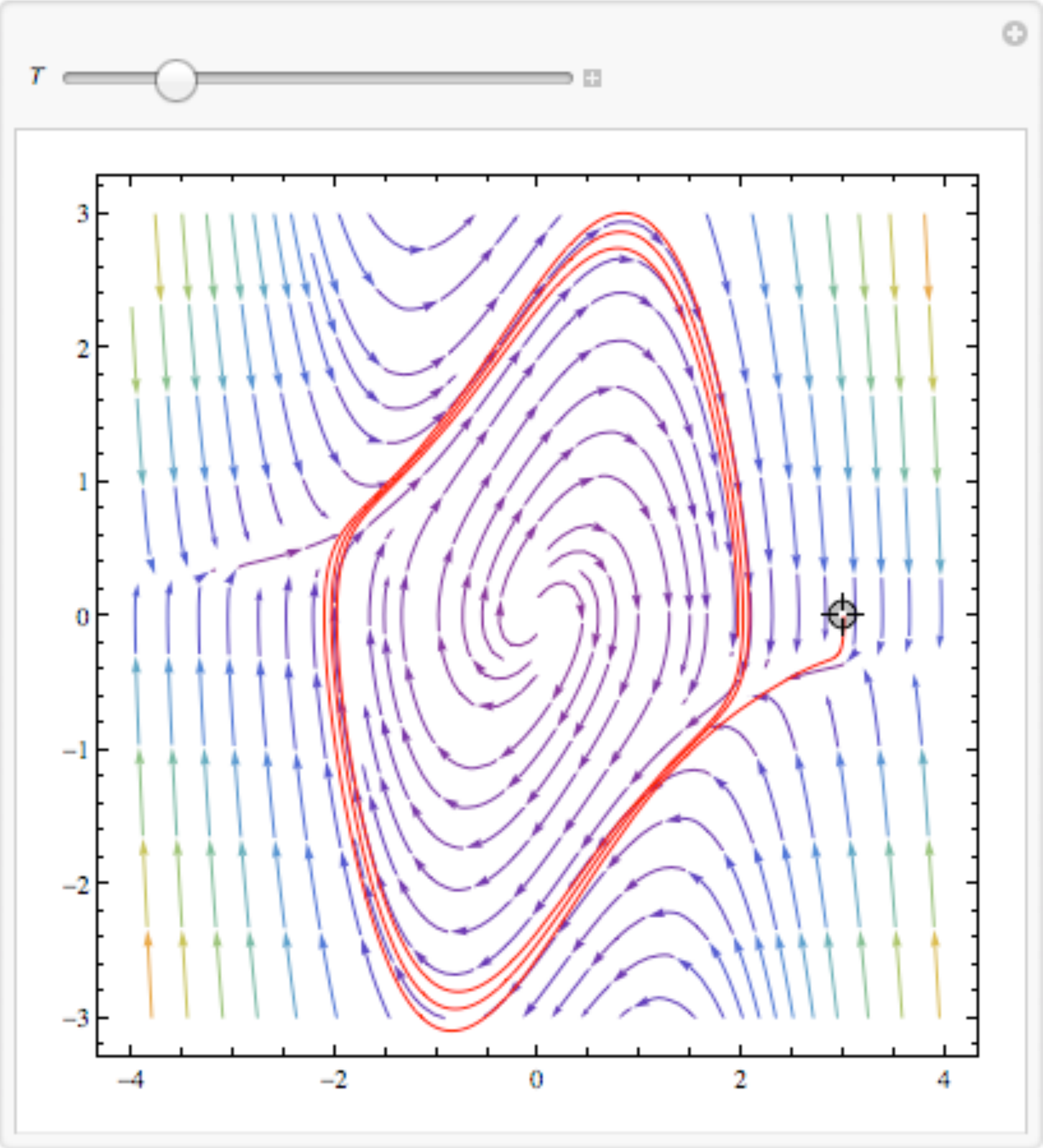
[Steven Strogatz, *Nonlinear dynamics and chaos*]

[Vyas et al., *Computation through neural population dynamics*, 2020]

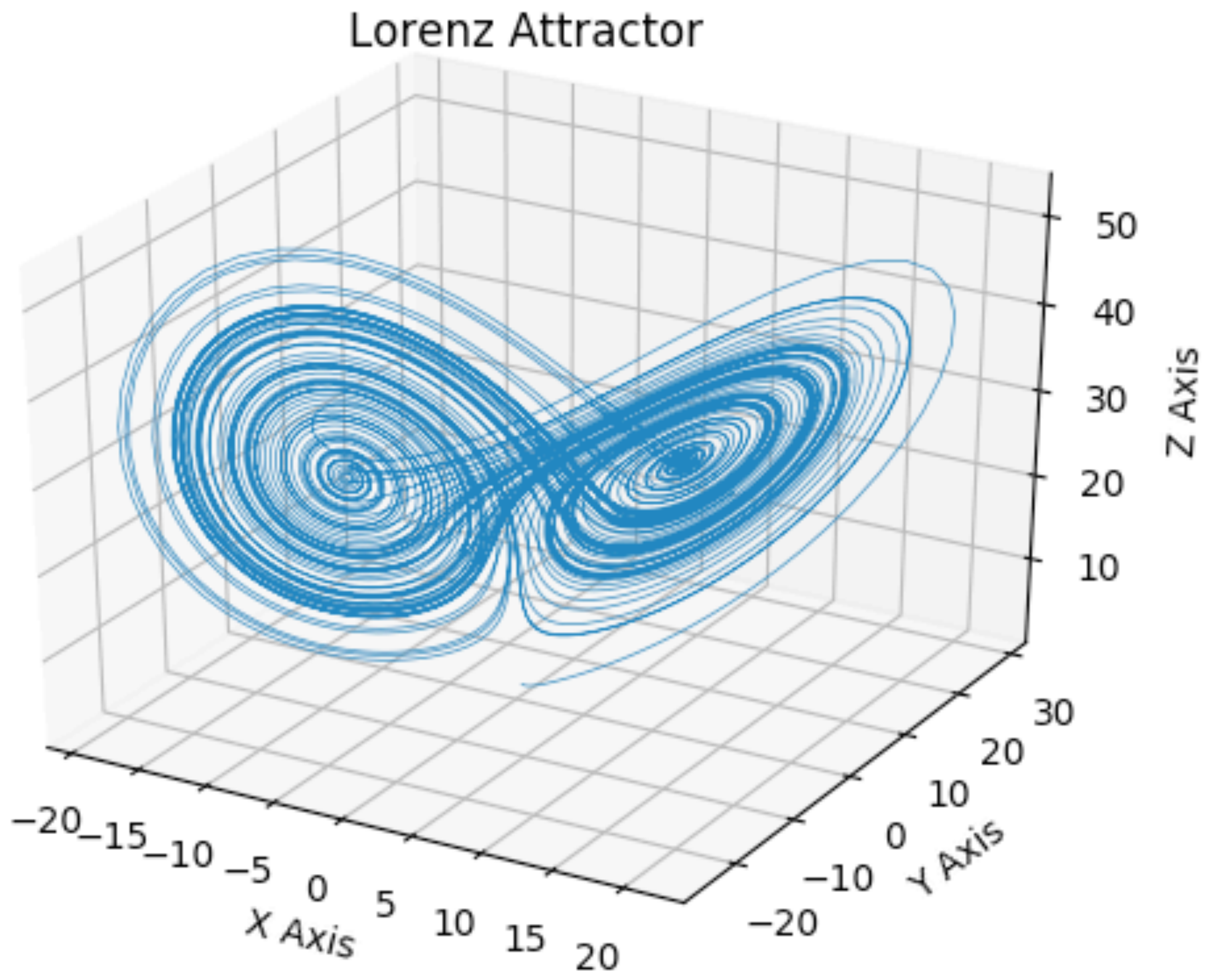
Fixed points



Periodic behavior



Chaotic behavior

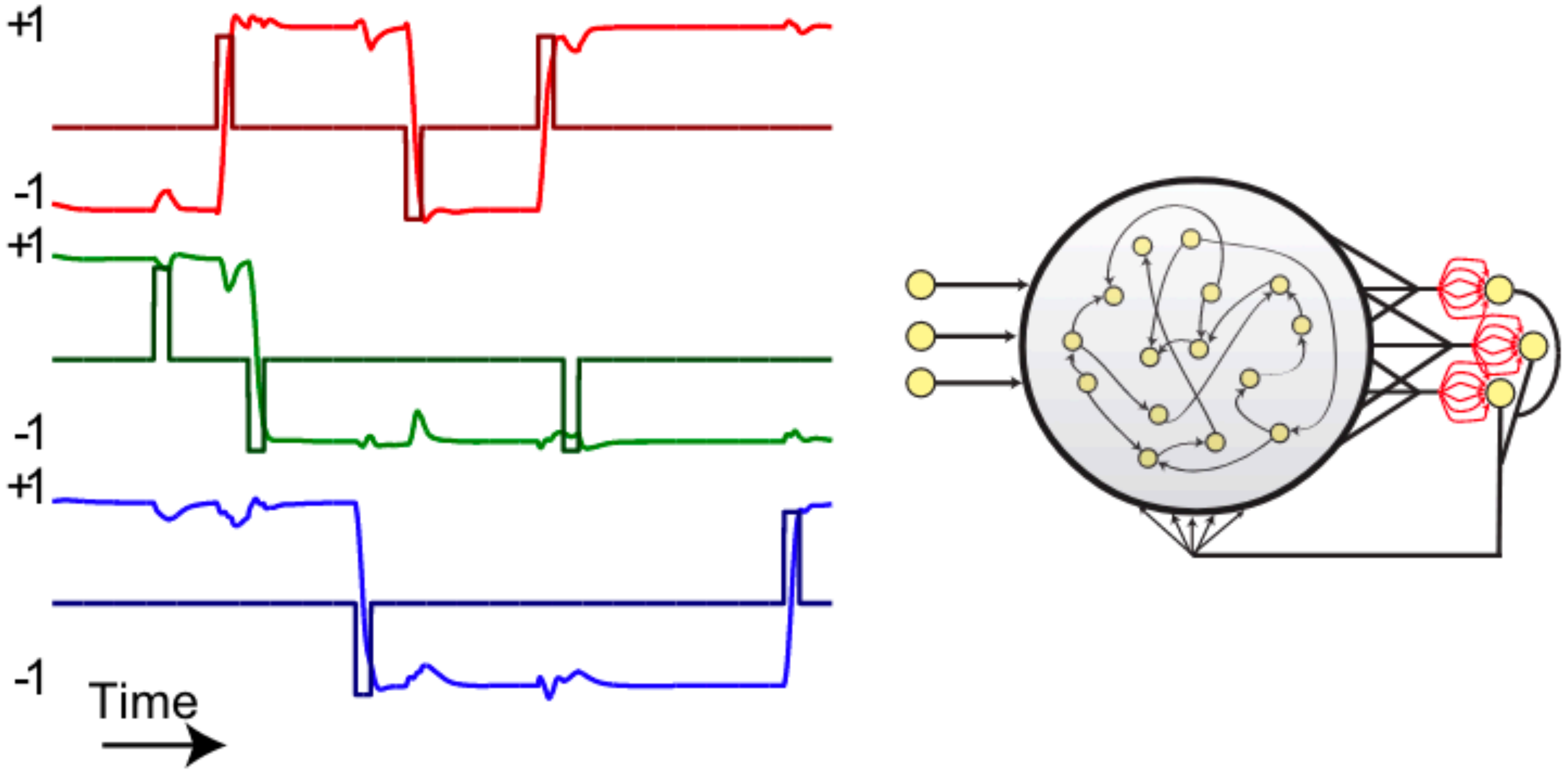


> 3 dimensions

Reverse-engineering RNNs

[Barak & Sussillo, *Opening the black-box: low-dimensional dynamics in high-dimensional RNNs*, 2013]

Flip-flop task:



Reverse-engineering RNNs

[Barak & Sussillo, *Opening the black-box: low-dimensional dynamics in high-dimensional RNNs*, 2013]

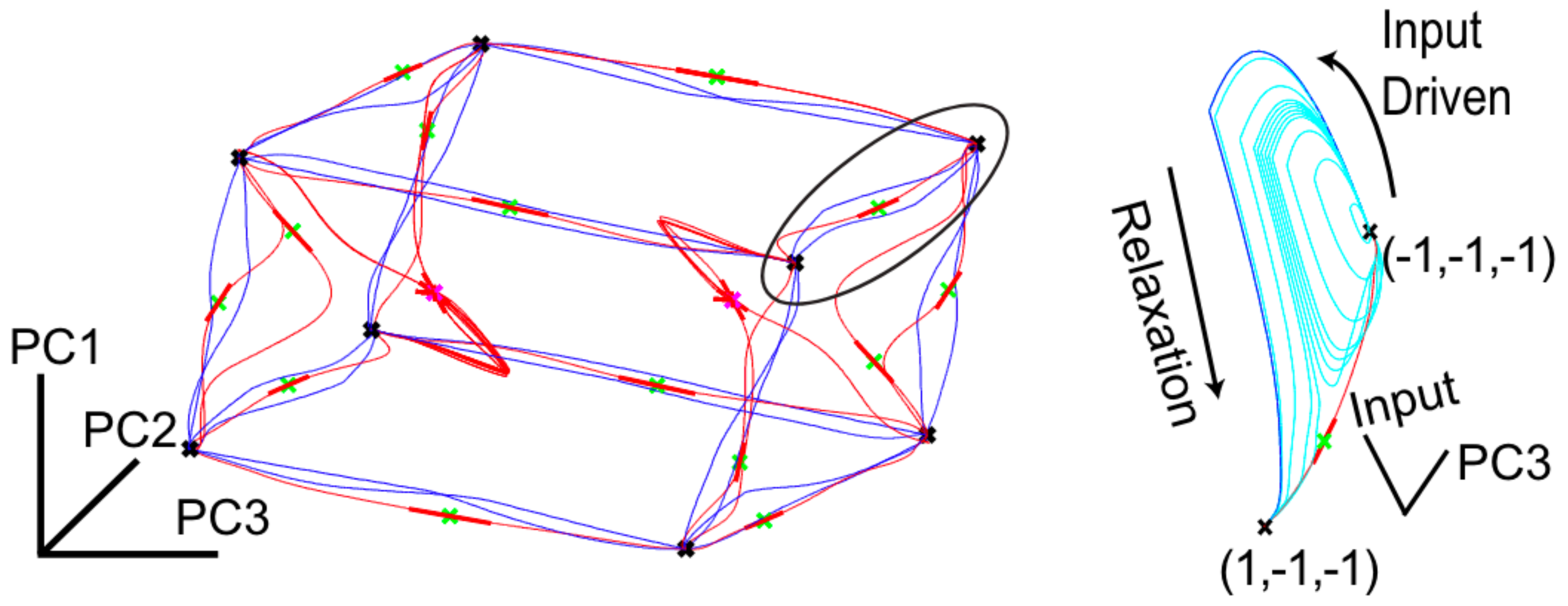
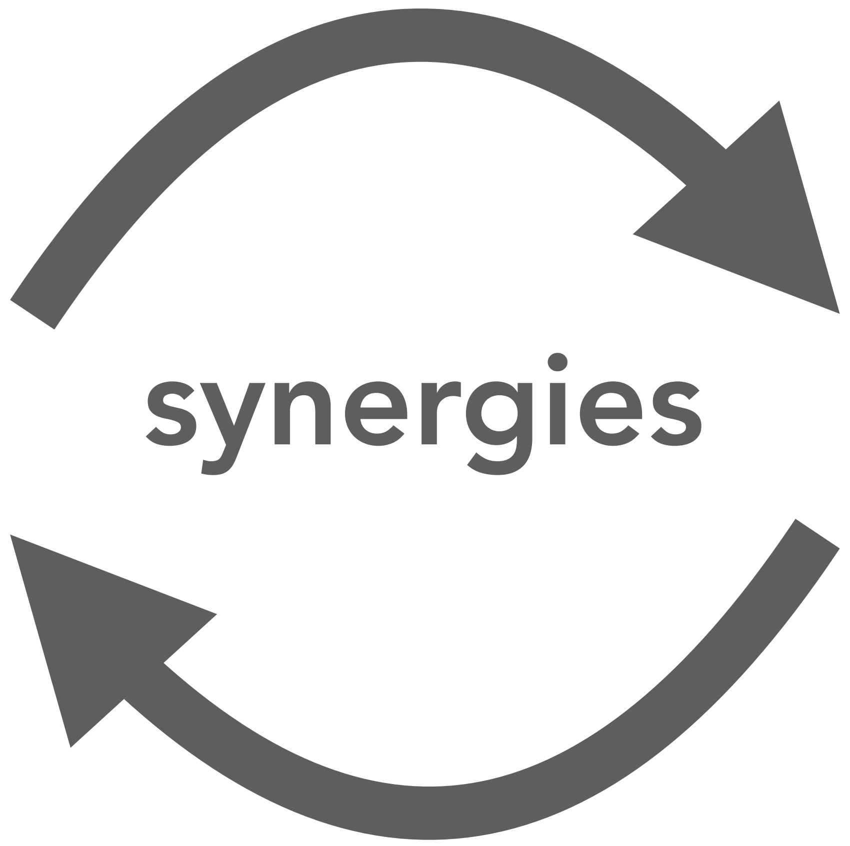


Figure 3: **Low-dimensional phase space representation of 3-bit flip-flop task.**

To summarize

BRAINS **ANNs**

- distributed computation machines
- use low-dimensional dynamics
- exhibit hierarchical computations

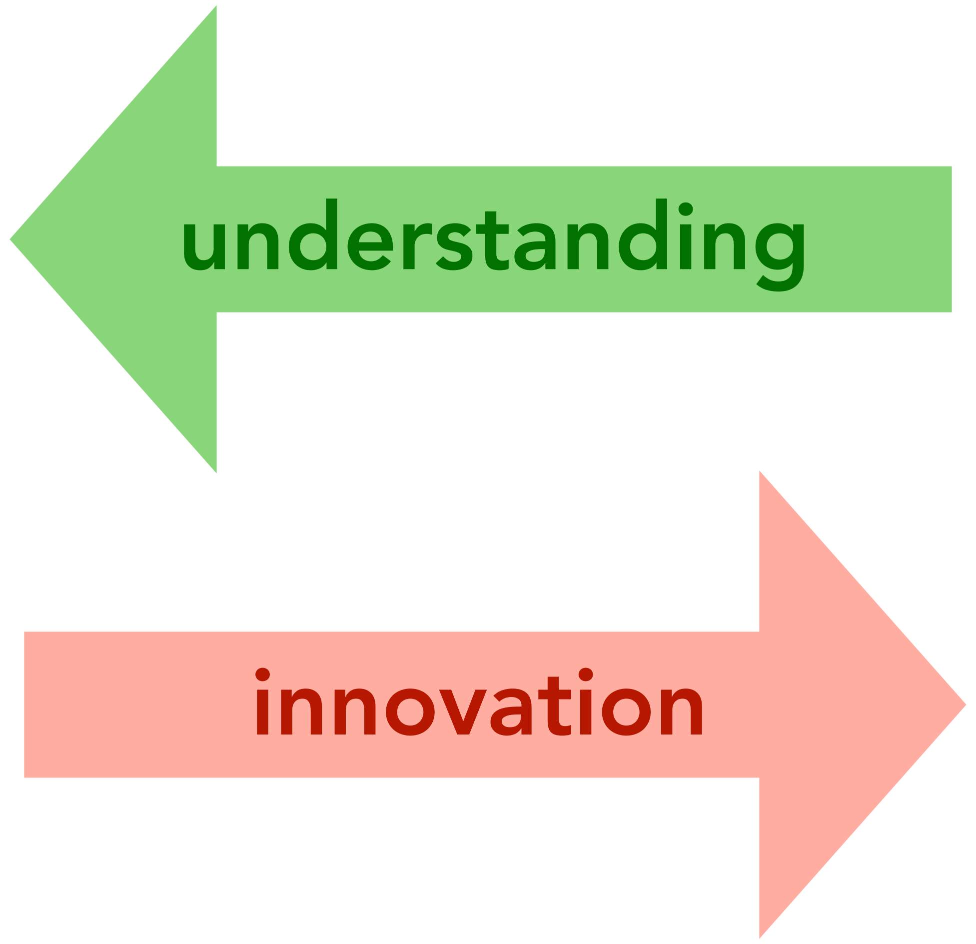


To summarize

BRAINS **ANNs**

- distributed computation machines
- use low-dimensional dynamics
- exhibit hierarchical computations

Still more powerful



Much simpler to analyse!

Possible insights

- *Which architecture leads to more similarity with brain recordings ?*
- *How does structure/connectivity relate to computations ?*
- *Are biological constraints a consequence of computational requirements ?*
- *Are some learning algorithms more similar to biological observations ?*

Underlying arguments

In favor

- If you want to understand the brain, start by (supposedly much simpler) ANNs !
- Cognition should be understood at the level of networks, not single neurons
- No tedious handcrafting needed

Critiques

- Learning is completely different
- ANNs live in a “tiny world” and optimize very specific objectives

Questions ?