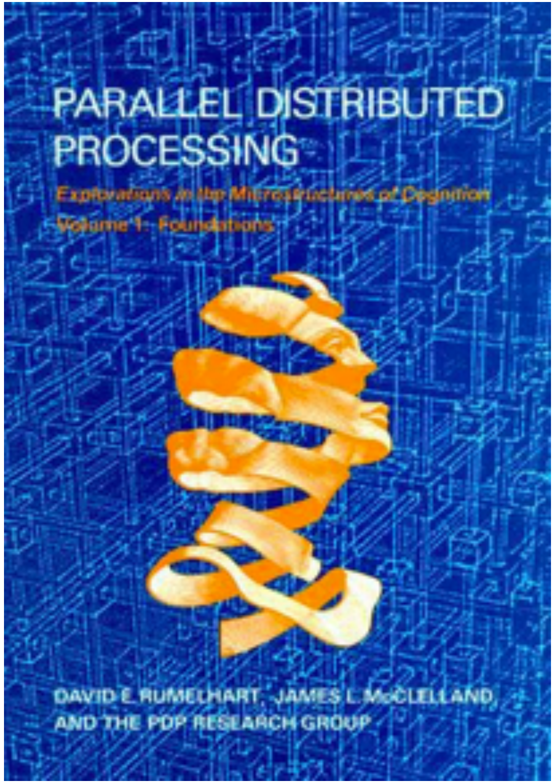
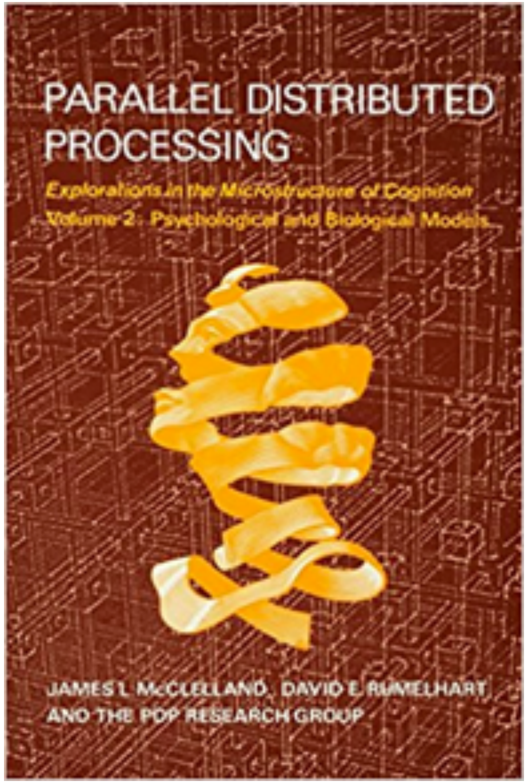
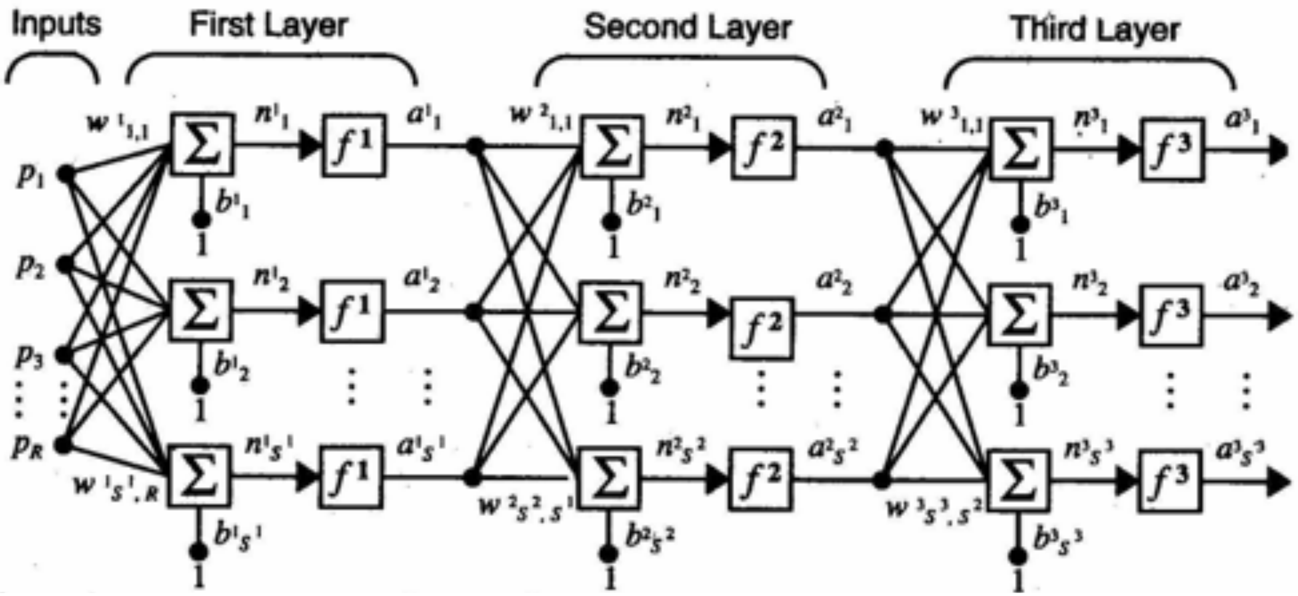
Deep Reinforcement Learning

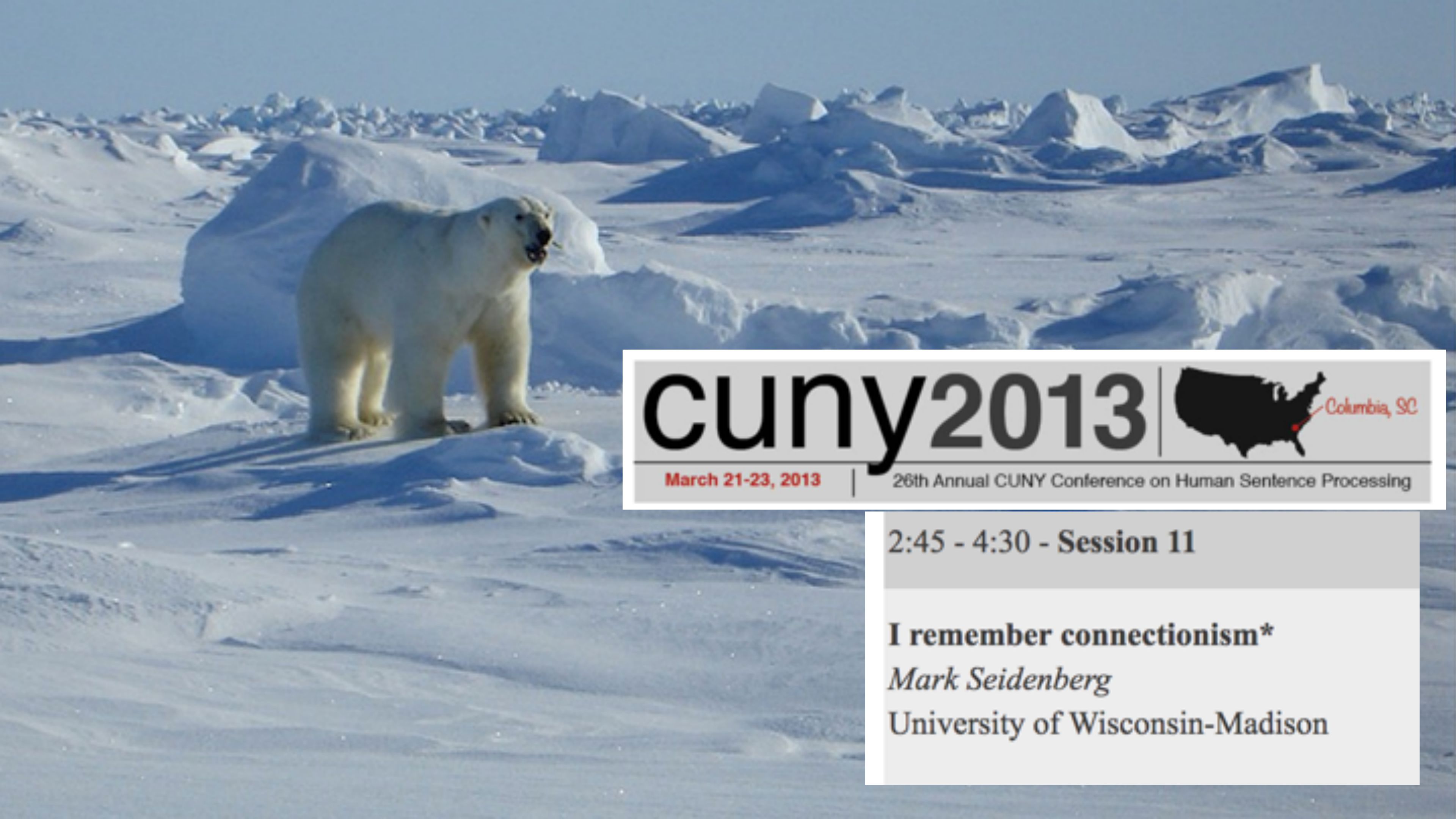
Matthew Botvinick
DeepMind, London UK
Gatsby Computational Neuroscience Unit, UCL

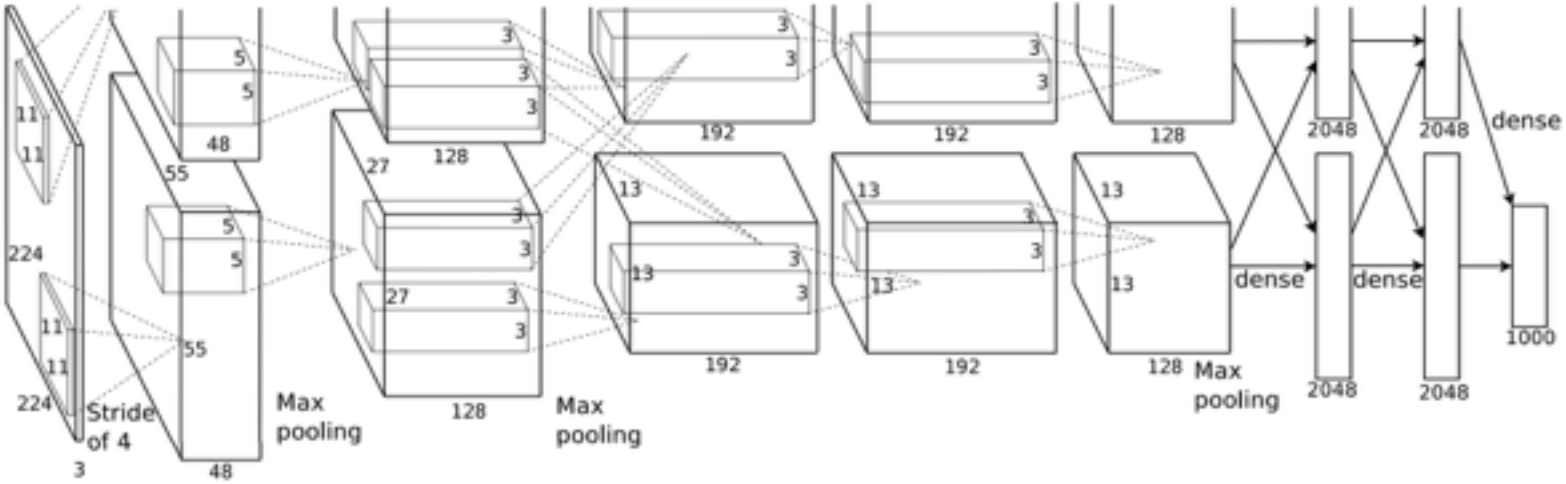

PARALLEL DISTRIBUTED PROCESSING
Explorations in the Microstructures of Cognition
Volume 1: Foundations
DAVID E. RUMELHART, JAMES L. McCLELLAND,
AND THE PDP RESEARCH GROUP

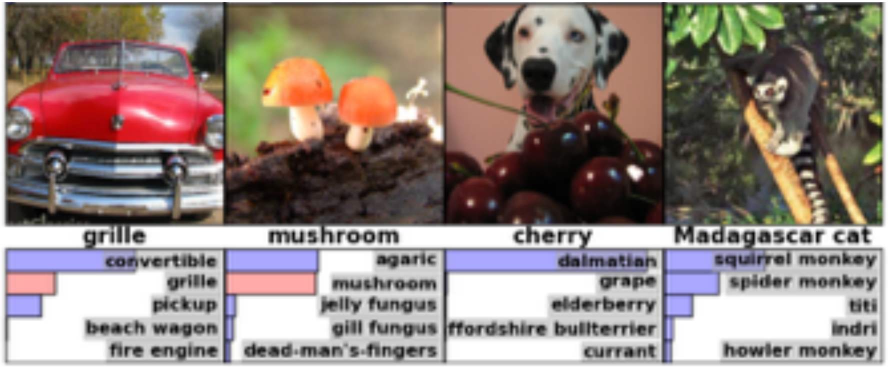

PARALLEL DISTRIBUTED PROCESSING
Explorations in the Microstructure of Cognition
Volume 2: Psychological and Biological Models
JAMES L. McCLELLAND, DAVID E. RUMELHART,
AND THE PDP RESEARCH GROUP



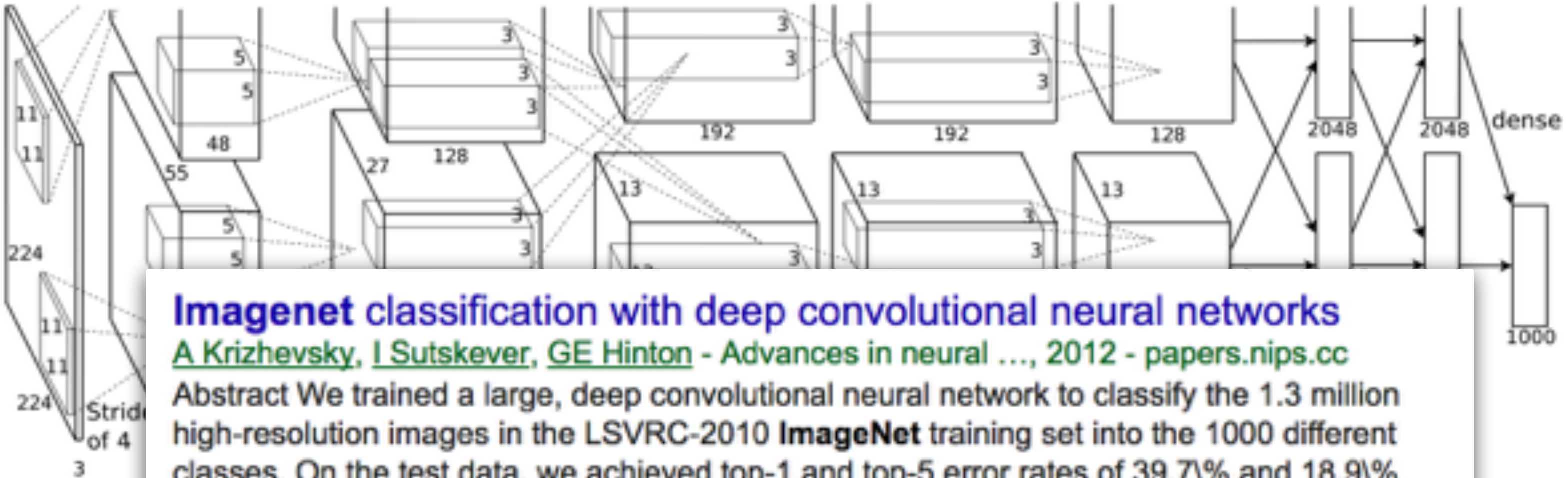
Hagan, 1996


cuny2013
Columbia, SC
March 21-23, 2013
26th Annual CUNY Conference on Human Sentence Processing
2:45 - 4:30 - Session 11
I remember connectionism*
Mark Seidenberg
University of Wisconsin-Madison





Krizhevsky, Sutskever & Hinton (2012)

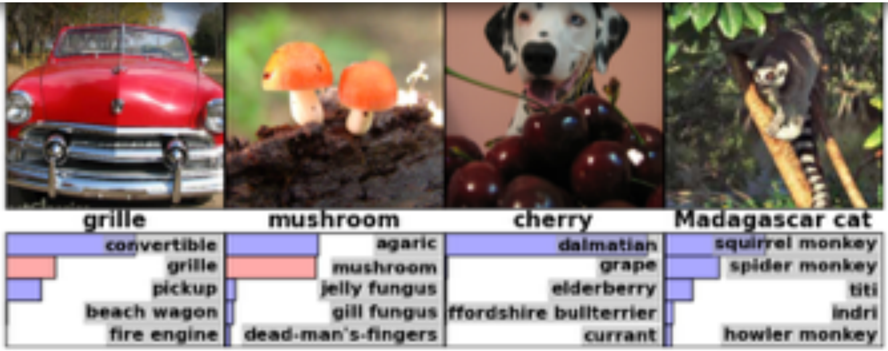


Imagenet classification with deep convolutional neural networks

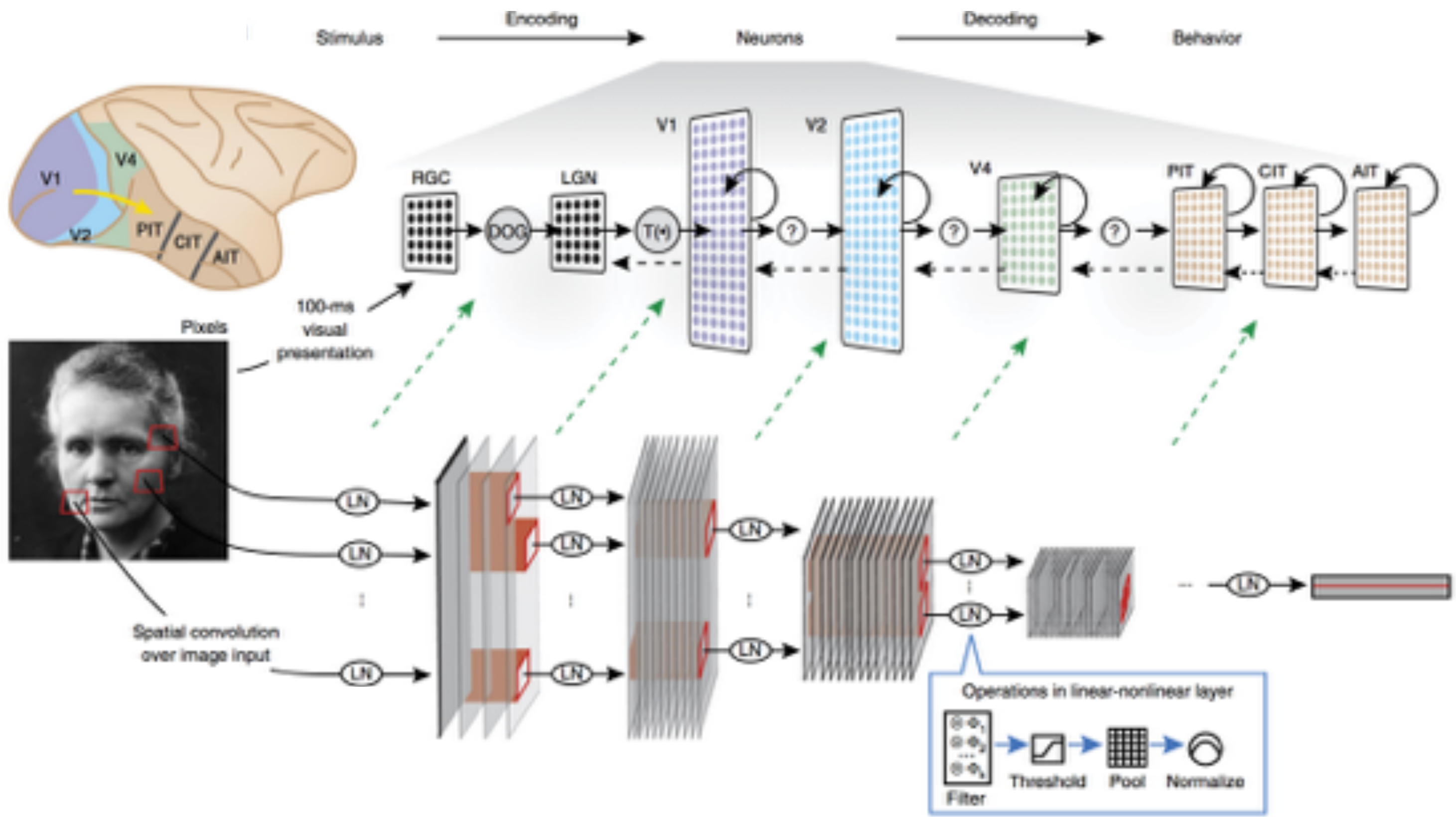
A Krizhevsky, I Sutskever, GE Hinton - Advances in neural ..., 2012 - papers.nips.cc

Abstract We trained a large, deep convolutional neural network to classify the 1.3 million high-resolution images in the LSVRC-2010 **ImageNet** training set into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 39.7\% and 18.9\%

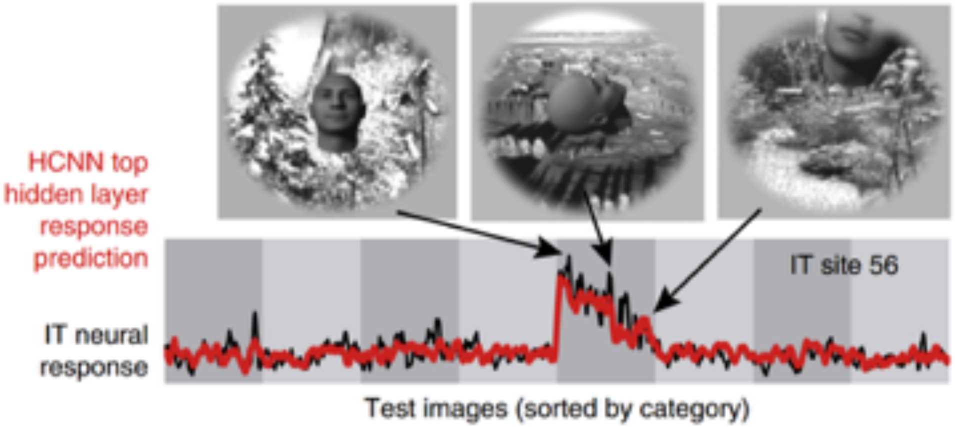
Cited by 12388 Related articles All 97 versions Cite Save

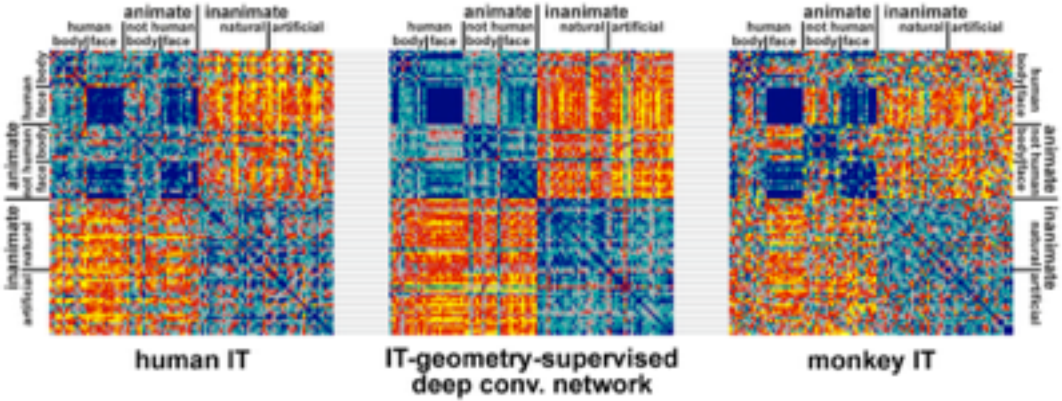


Krizhevsky, Sutskever & Hinton (2012)

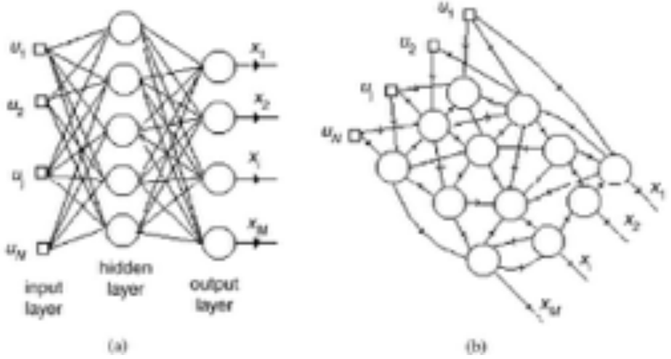


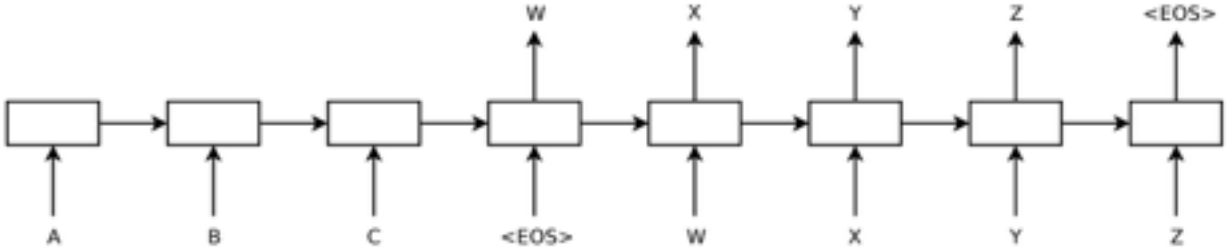
Yamins & DiCarlo, 2016



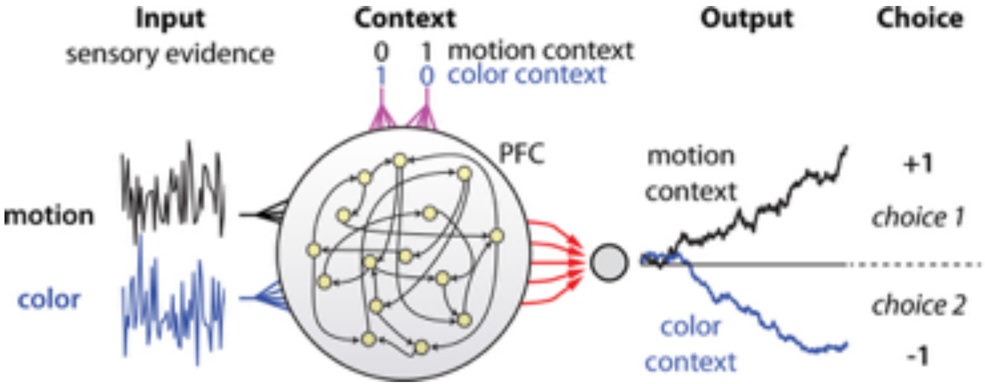


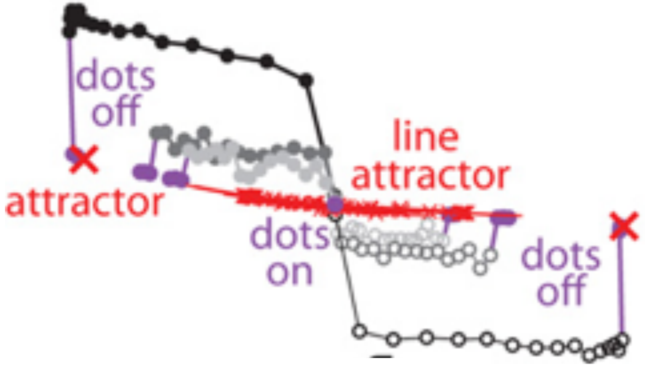
Yamins & DiCarlo, 2016; Khaligh-Razavi & Kriegeskorte, 2014; Kietzmann et al., 2017





Sutskever et al., 2014





Mante et al., *Nature*, 2013

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Scientific reports, 6.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nature Neuroscience, 19(3), 356–365.

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. Journal of Neuroscience, 35(27), 10005-10014.

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Computational Biology, 10(11), 1–29.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences, 111(23), 8619-8624.

Yamins, D. L., Hong, H., Cadieu, C., & DiCarlo, J. J. (2013). Hierarchical modular optimization of convolutional networks achieves representations similar to macaque IT and human ventral stream. In Advances in neural information processing systems (pp. 3093-3101).

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature, 503(7474), 78-84.

Towards a Biologically Plausible Backprop

Benjamin Scellier and Yoshua Bengio*
Université de Montréal, Montreal Institute for Learning Algorithms

September 21, 2016

This work contributes several new elem
brains. We introduce a very general and ab
defined implicitly through an energy functi
for the first phase (when the prediction is m
Hebbian learning algorithm in the continu
computational graphs (i.e. standard backp
framework. One advantage of our framewo
only nudging the first-phase fixed point tow
supervised neural network, the output unit
the output layer propagates backward in th
contains information about the error deriva
descent with respect to an objective functio

1 Introduction

We introduce a very general and abstract fr
the data and the parameters of the model (
descent on an objective function. We also p
But we expect that the general framework c
In section 3, we give a new descriptio
our general framework. In the model, inp
a local minimum of an energy function. T
corresponds to the first phase of the algorit

Overcoming catastrophic forgetting in neural networks

James Kirkpatrick[a,1], Razvan Pascanu[a], Neil Rabinowitz[a], Joel Veness[a], Guillaume Desjardins[a], Andrei A. Rusu[a], Kieran Milan[a], John Quan[a], Tiago Ramalho[a], Agnieszka Grabska-Barwinska[a], Demis Hassabis[a], Claudia Clopath[b], Dharshan Kumaran[a], and Raia Hadsell[a]

[a]DeepMind, London EC4 5TW, United Kingdom; and [b]Bioengineering Department, Imperial College London, London SW7 2AZ, United Kingdom

Edited by James L. McClelland, Stanford University, Stanford, CA, and approved February 13, 2017 (received for review July 19, 2016)

In marked contrast to artificial neural networks, humans
r to be able to learn in a continual
dence suggests that the mammalian
hic forgetting by protecting previously
eocortical circuits (11–14). When a
ll, a proportion of excitatory synapses
anifests as an increase in the volume
ines of neurons (13). Critically, these
persist despite the subsequent learning
for retention of performance several
these spines are selectively "erased,"
orgotten (11, 12). This provides causal
hanisms supporting the protection of
es are critical to retention of task per-
ental findings—together with neurobi-
cascade model (15, 16)—suggest that
eocortex relies on task-specific synap-
knowledge is durably encoded by ren-
apses less plastic and therefore stable

trate that task-specific synaptic consol-
lution to the continual-learning prob-
nce. We develop an algorithm anal-
dation for artificial neural networks,
tic weight consolidation (EWC). This
ning on certain weights based on how
iously seen tasks. We show how EWC
d learning and reinforcement learn-
ral tasks sequentially without forget-
d contrast to previous deep-learning

nature COMMUNICATIONS

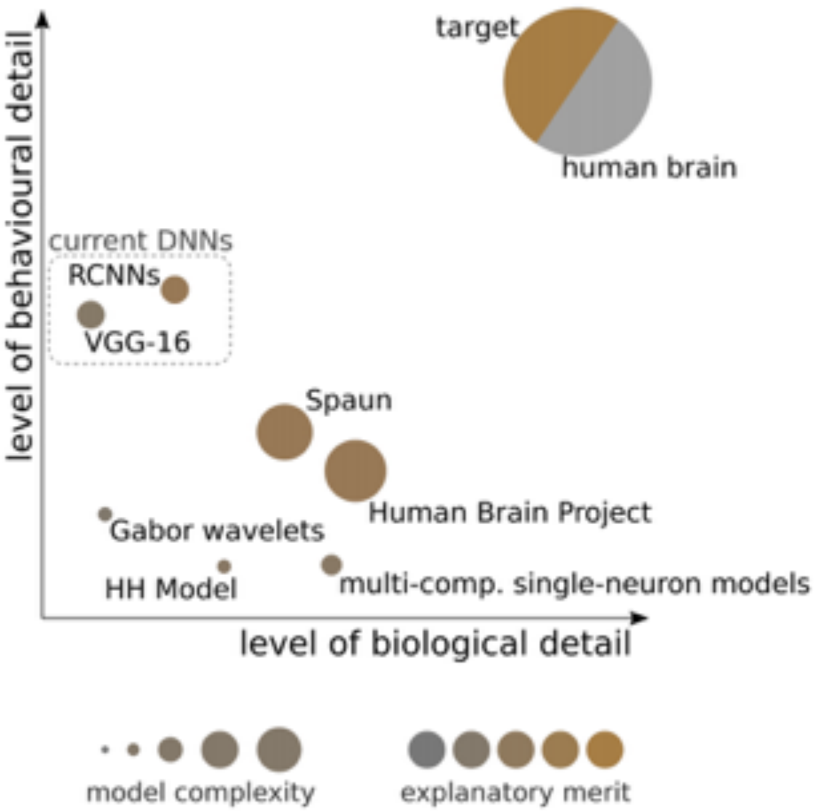
ARTICLE

Received 7 Jan 2016 | Accepted 16 Sep 2016 | Published 8 Nov 2016 DOI: 10.1038/ncomms13276 OPEN

Random synaptic feedback weights support error backpropagation for deep learning

Timothy P. Lillicrap[1,2], Daniel Cownden[3], Douglas B. Tweed[4,5] & Colin J. Akerman[1]

The brain processes information through multiple layers of neurons. This deep architecture is



Kietzmann et al., *bioRxiv*, 2017

“Reinforcement learning is learning what to do — how to map situations to actions — so as to maximize a numerical reward signal. The learner is not told which actions to take...but instead must discover which actions yield the most reward by trying them.”

— Sutton & Barto, 1998

Return: $$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

Value Functions:

$$Q^\pi(s,a) = E_\pi\{R_t | s_t = s, a_t = a\} = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \,\middle|\, s_t = s, a_t = a\right\}$$

Temporal Difference Learning:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\right]$$

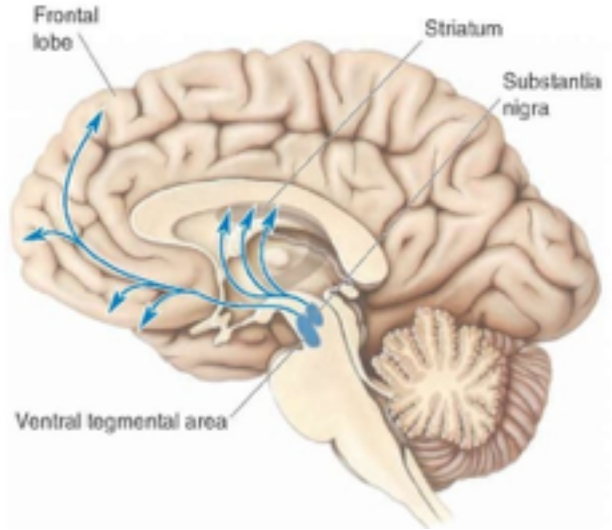
Return: $$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

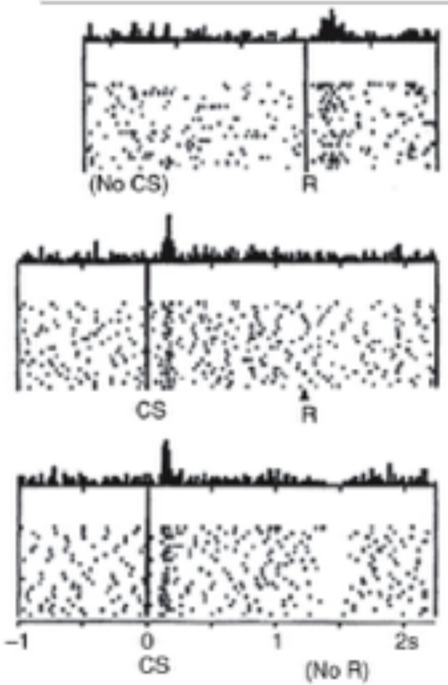
Value Functions:

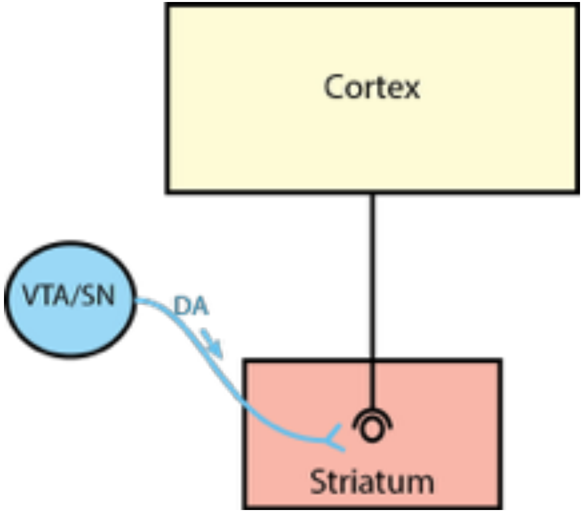
$$Q^\pi(s,a) = E_\pi\{R_t | s_t = s, a_t = a\} = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \,\middle|\, s_t = s, a_t = a\right\}$$

Temporal Difference Learning:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[\underline{r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)}\right]$$

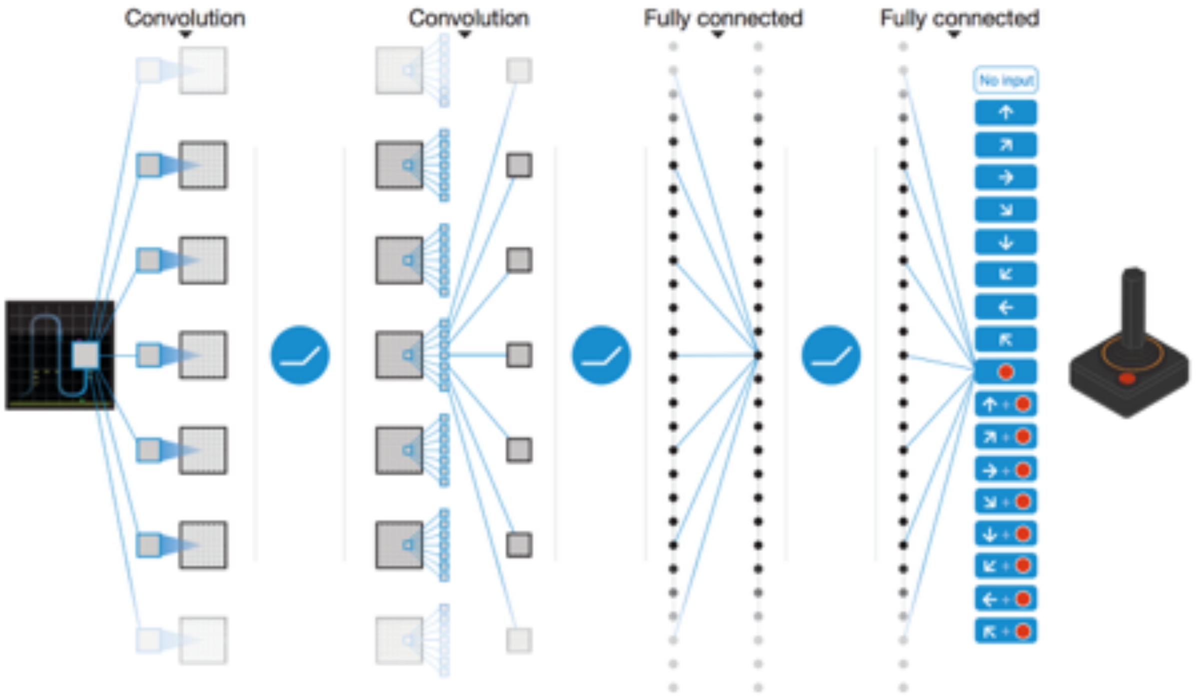




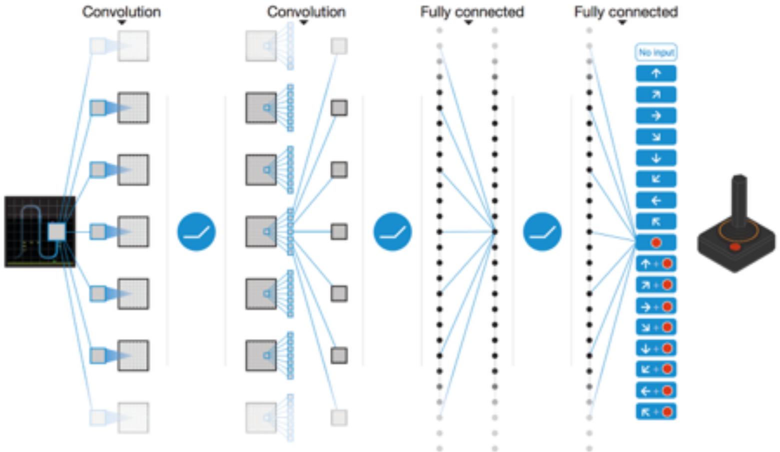


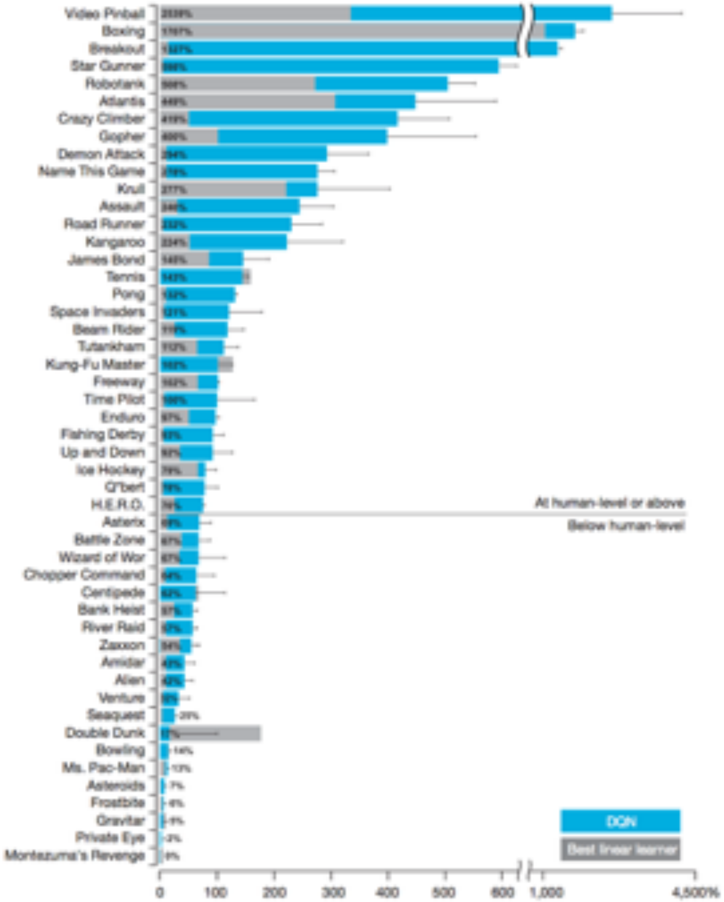
Schultz et al, Science (1997)

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

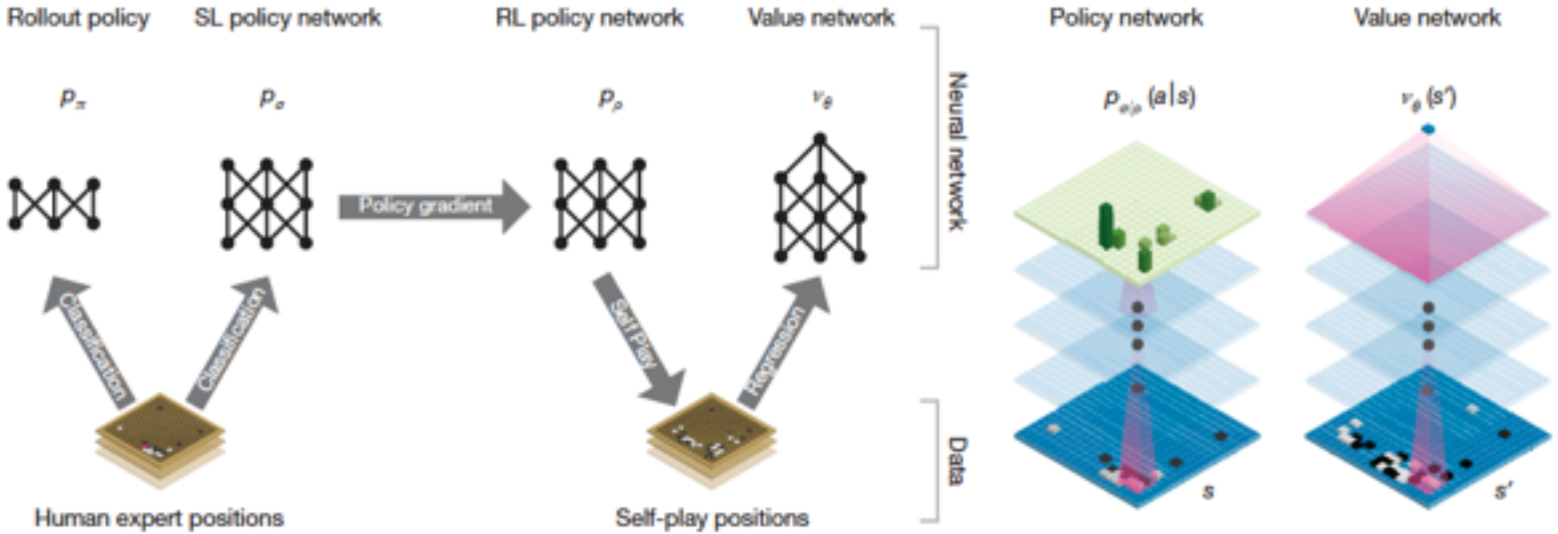


Mnih *et al, Nature* (2015)



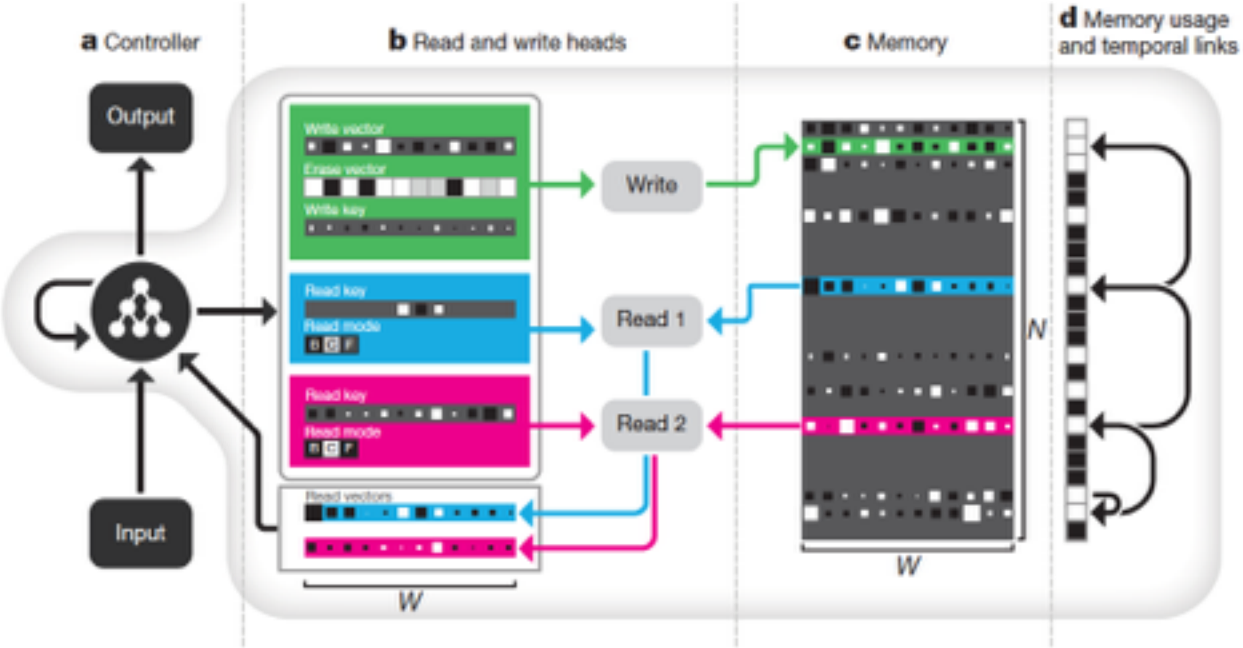


Mnih *et al, Nature* (2015)

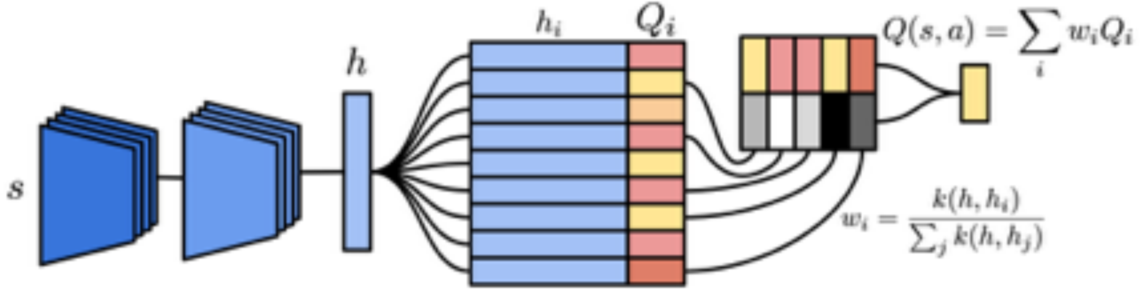


Silver *et al, Nature* (2015)

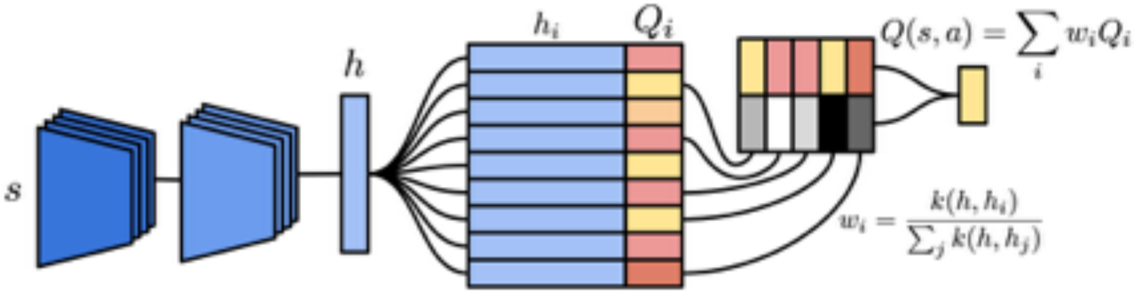
Mastering the game of **Go** with deep neural networks and tree search
D **Silver**, A Huang, CJ Maddison, A Guez, L Sifre… - **Nature**, 2016 - **nature**.com
The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. Here we introduce a new approach to computer Go that uses 'value
Cited by 1057 Related articles All 39 versions Cite Save

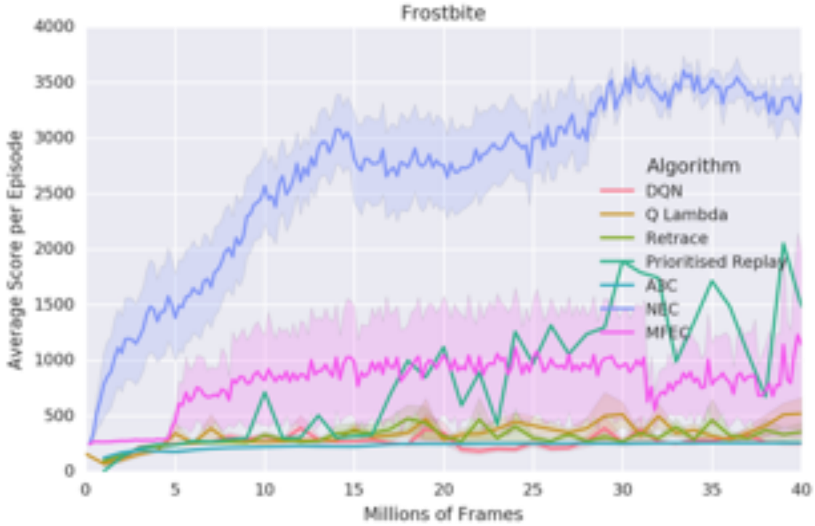


Graves *et al*, *Nature* (2016)



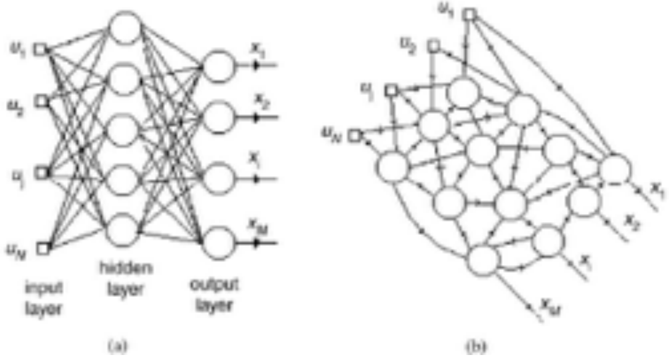
Pritzl *et al, arXiv* (2017)

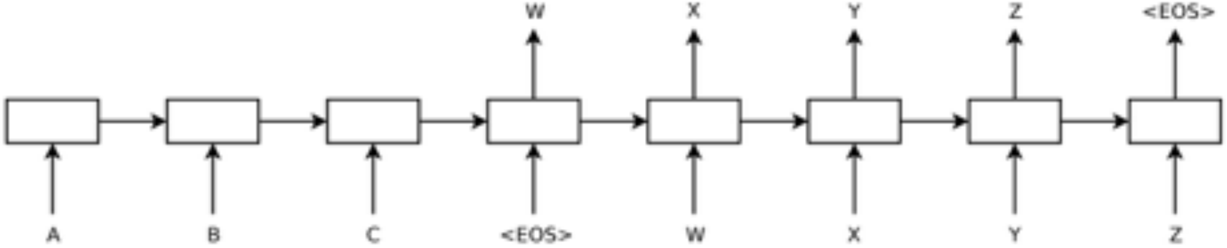




Pritzl *et al, arXiv* (2017)

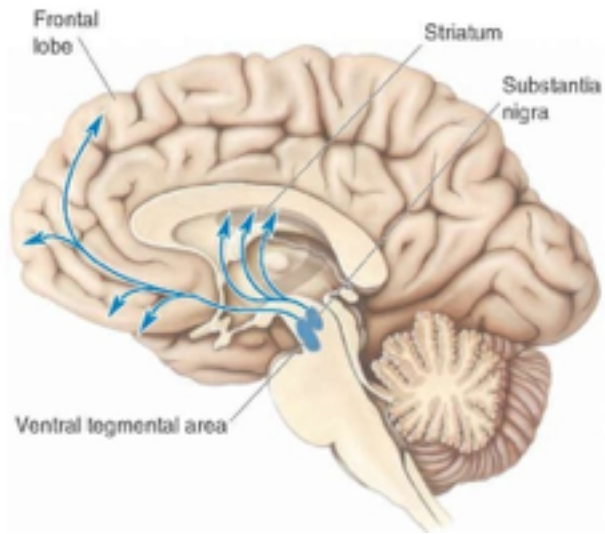
Pritzl *et al, arXiv* (2017)

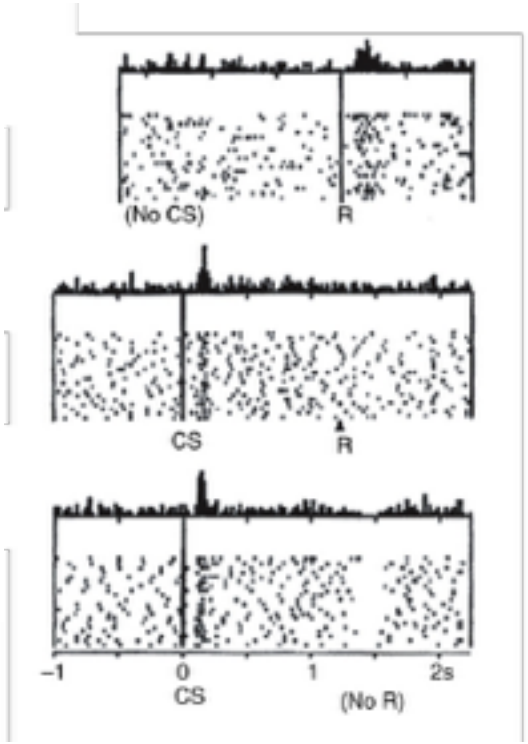


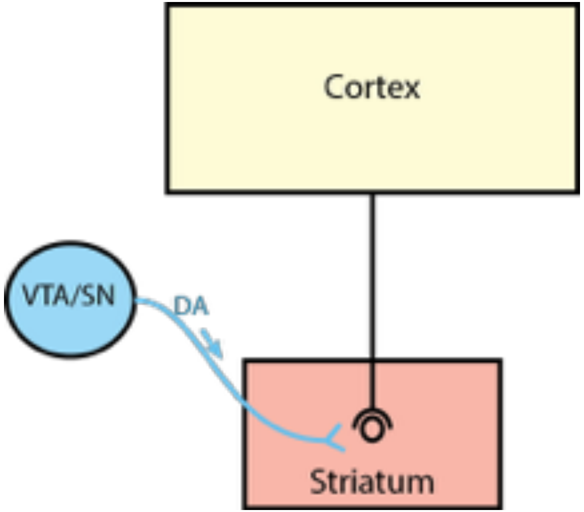


Sutskever et al., 2014

Jederberg et al., 2016







Schultz et al, Science (1997)

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Behavioral Sciences

Reinforcement learning, efficient coding, and the statistics of natural tasks

Matthew Botvinick[1,2,3], Ari Weinstein[1,2], Alec Solway[4] and Andrew Barto[5]

The application of ideas from computational reinforcement
learning has recently enabled dramatic advances in behavio
and neuroscientific research. For the most part, these
advances have involved insights concerning the algorithms
underlying learning and decision making. In the present artic
we call attention to the equally important but relatively
neglected question of how problems in learning and decis
making are internally represented. To articulate the significan
of representation for reinforcement learning we draw on th
concept of efficient coding, as developed in perception
research. The resulting perspective exposes a range of no
goals for behavioral and neuroscientific research, highlight
in particular the need for research into the statistical structure
naturalistic tasks.

Addresses
[1] Princeton Neuroscience Institute, Princeton University, Princeton,
United States
[2] Department of Psychology, Princeton University, Princeton, NJ, U
States
[3] Google DeepMind, London, United Kingdom
[4] Virginia Tech Carilion Research Institute, Roanoke, VA, United St
[5] College of Information and Computer Sciences, University of
Massachusetts, Amherst, MA, United States

Corresponding author: Botvinick, Matthew (matthewb@princeton.

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Behavioral Sciences

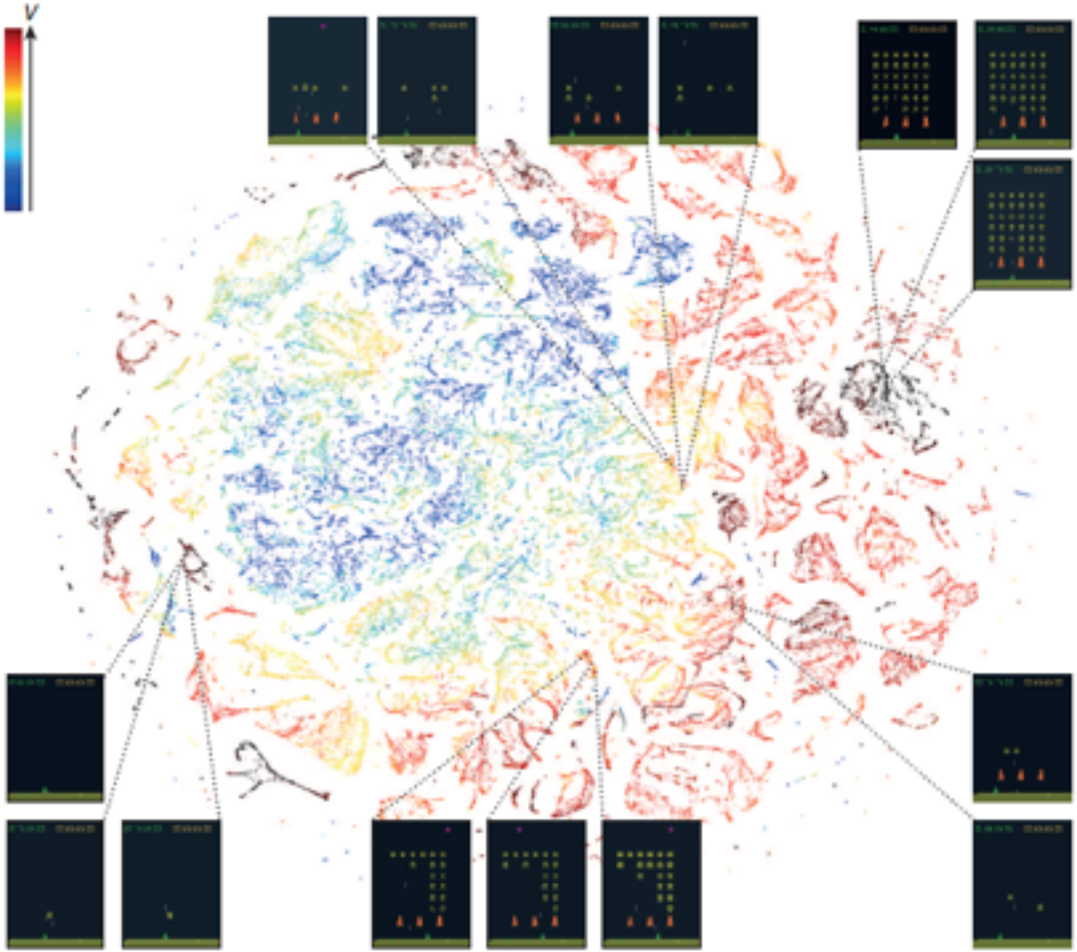
Discovering latent causes in reinforcement learning

Samuel J Gershman[1], Kenneth A Norman[2] and Yael Niv[2]

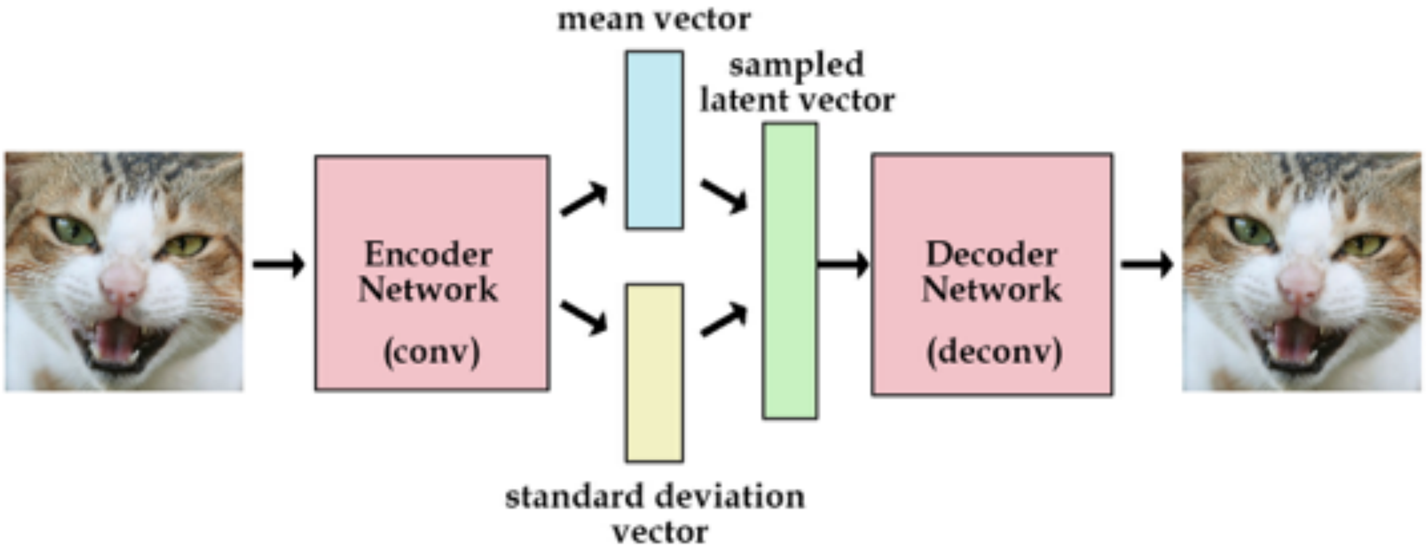
Effective reinforcement learning hinges on having an appropriate state representation. But where does this representation come from? We argue that the brain discovers state representations by trying to infer the latent causal structure of the task at hand, and assigning each latent cause to a separate state. In this paper, we review several implications of this latent cause framework, with a focus on Pavlovian conditioning. The framework suggests that conditioning is not the acquisition of associations between cues and outcomes, but rather the acquisition of associations between latent causes and observable stimuli. A latent cause interpretation of conditioning enables us to begin answering questions that have frustrated classical theories: Why do extinguished responses sometimes return? Why do stimuli presented in compound sometimes summate and sometimes do not? Beyond conditioning, the principles of latent causal inference may provide a general theory of structure learning

extend a familiar category to incorporate this unusual example, or to postulate a novel category. We refer to the process of parsing experience into groups or delineating the boundaries of generalization among examples as *structure learning*. The idea is that this parsing process attempts to follow the true causal structure in the world: Experiences that are all instances of the same cause should be grouped together and generalized over, while experiences that are due to different underlying causes should be separated in our mind. Recent work has begun to unravel the cognitive and neural mechanisms supporting structure learning (see [2,3] for reviews). Our focus here is on the role of structure in reinforcement learning.

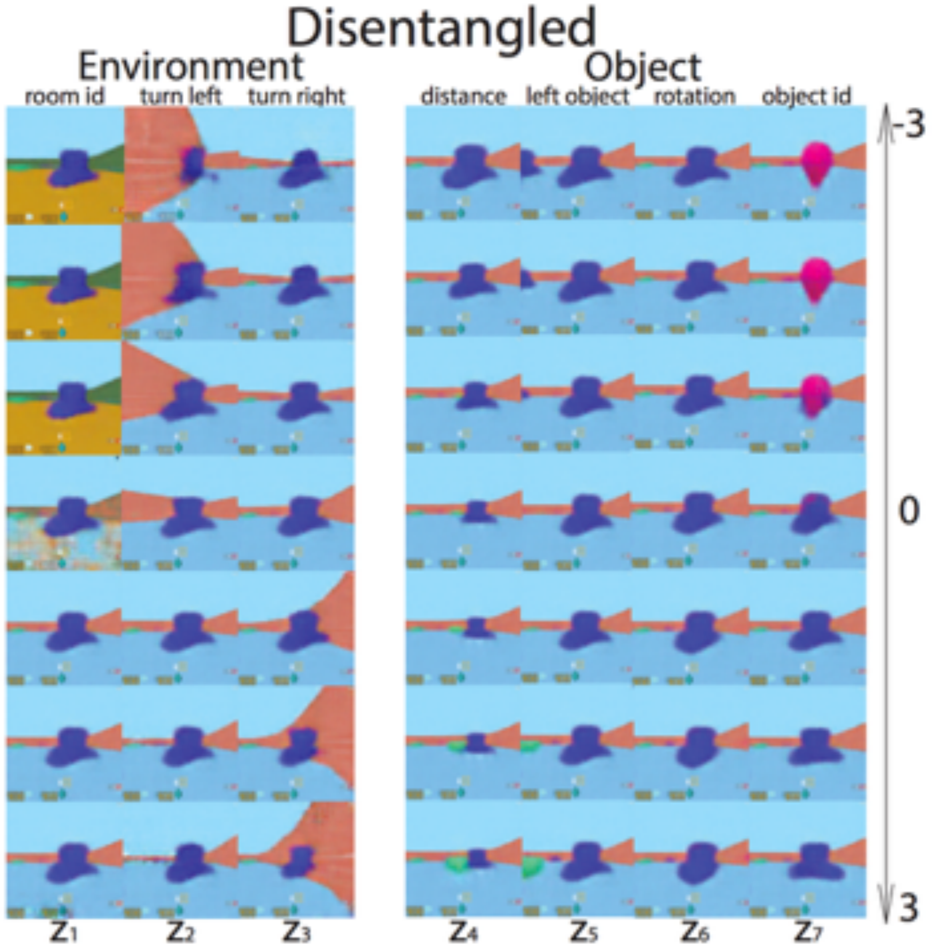
Structure learning is fundamental to reinforcement learning because these algorithms rely on a representation of

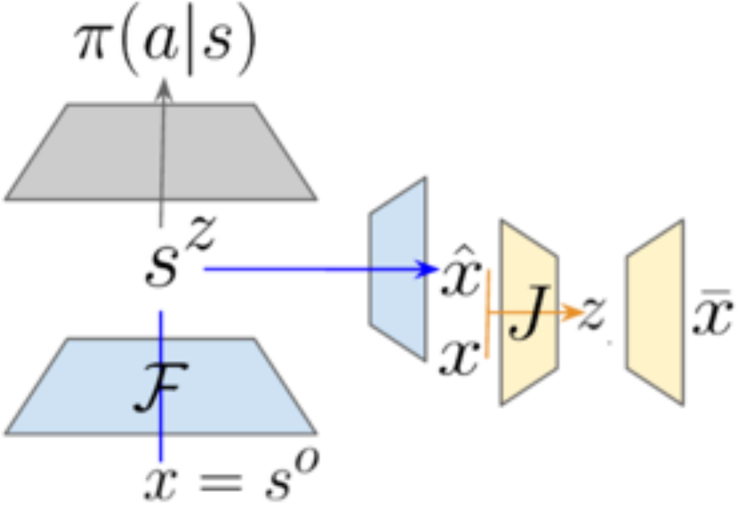


Mnih *et al, Nature* (2015)

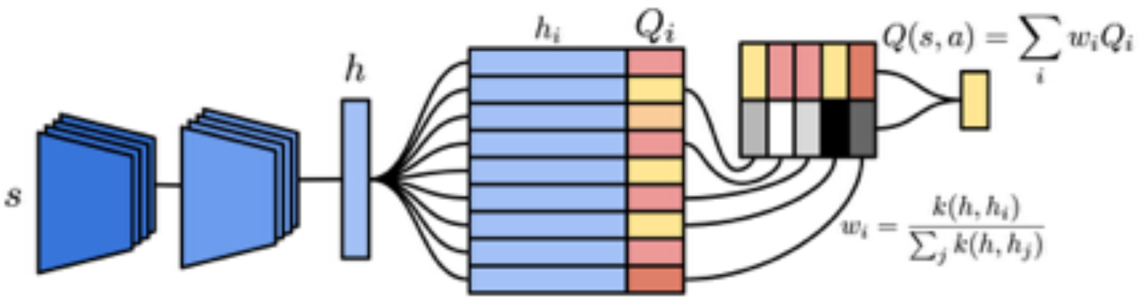

mean vector
sampled
latent vector
Encoder
Network
(conv)
Decoder
Network
(deconv)
standard deviation
vector

kvfrans.com





Higgins *et al.*, 2017



Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework

Samuel J. Gershman[1] and Nathaniel D. Daw[2]

[1]Department of Psychology and Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138; email: gershman@fas.harvard.edu

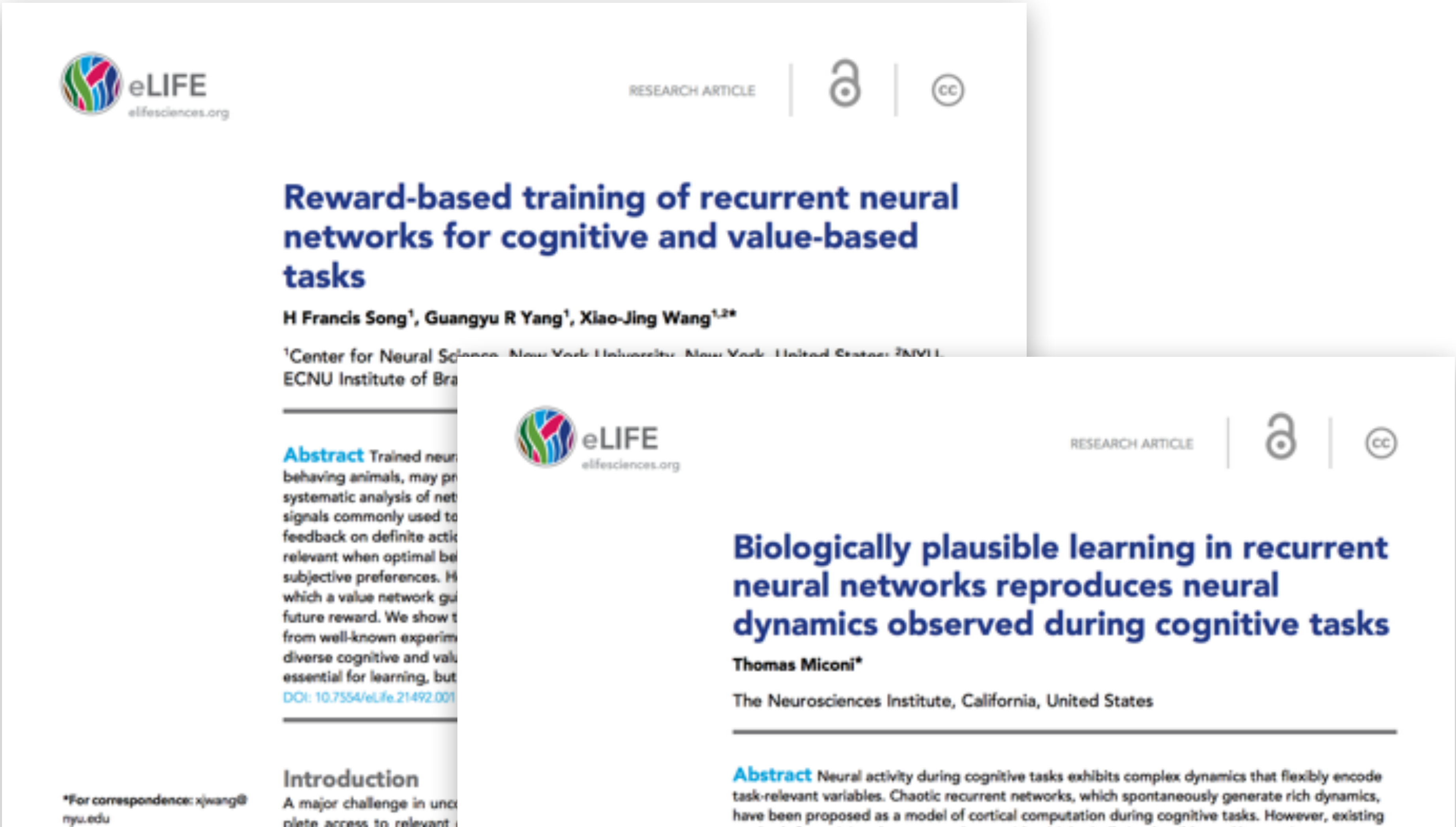
[2]Princeton Neuroscience Institute and Department of Psychology, Princeton University, Princeton, New Jersey 08544

Annu. Rev. Psychol. 2017. 68:101–28

First published online as a Review in Advance on

Keywords

reinforcement learning, memory, decision making



eLIFE
elifesciences.org

RESEARCH ARTICLE

Reward-based training of recurrent neural networks for cognitive and value-based tasks

H Francis Song[1], Guangyu R Yang[1], Xiao-Jing Wang[1,2]*

[1]Center for Neural Sc
ECNU Institute of Bra

Abstract Trained neur
behaving animals, may pr
systematic analysis of net
signals commonly used to
feedback on definite actio
relevant when optimal be
subjective preferences. H
which a value network gu
future reward. We show t
from well-known experim
diverse cognitive and valu
essential for learning, but
DOI: 10.7554/eLife.21492.001

Introduction

A major challenge in unc
plete access to relevant

*For correspondence: xjwang@nyu.edu

eLIFE
elifesciences.org

RESEARCH ARTICLE

Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks

Thomas Miconi*

The Neurosciences Institute, California, United States

Abstract Neural activity during cognitive tasks exhibits complex dynamics that flexibly encode task-relevant variables. Chaotic recurrent networks, which spontaneously generate rich dynamics, have been proposed as a model of cortical computation during cognitive tasks. However, existing

Thanks!