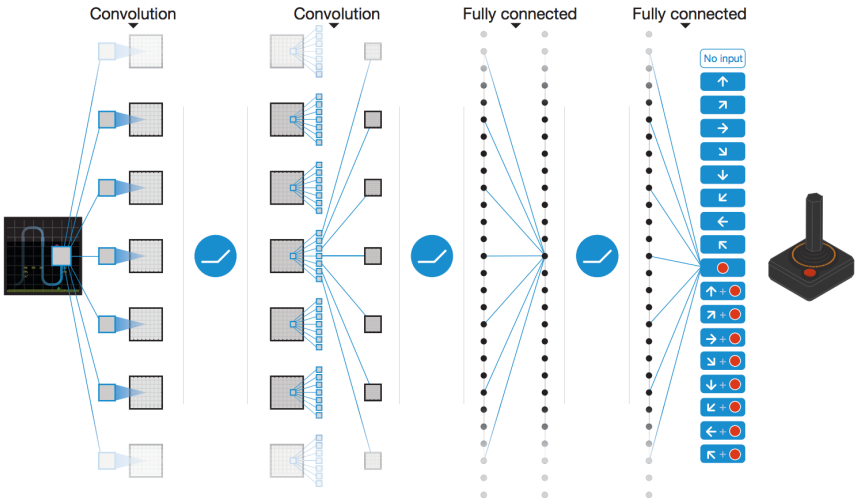
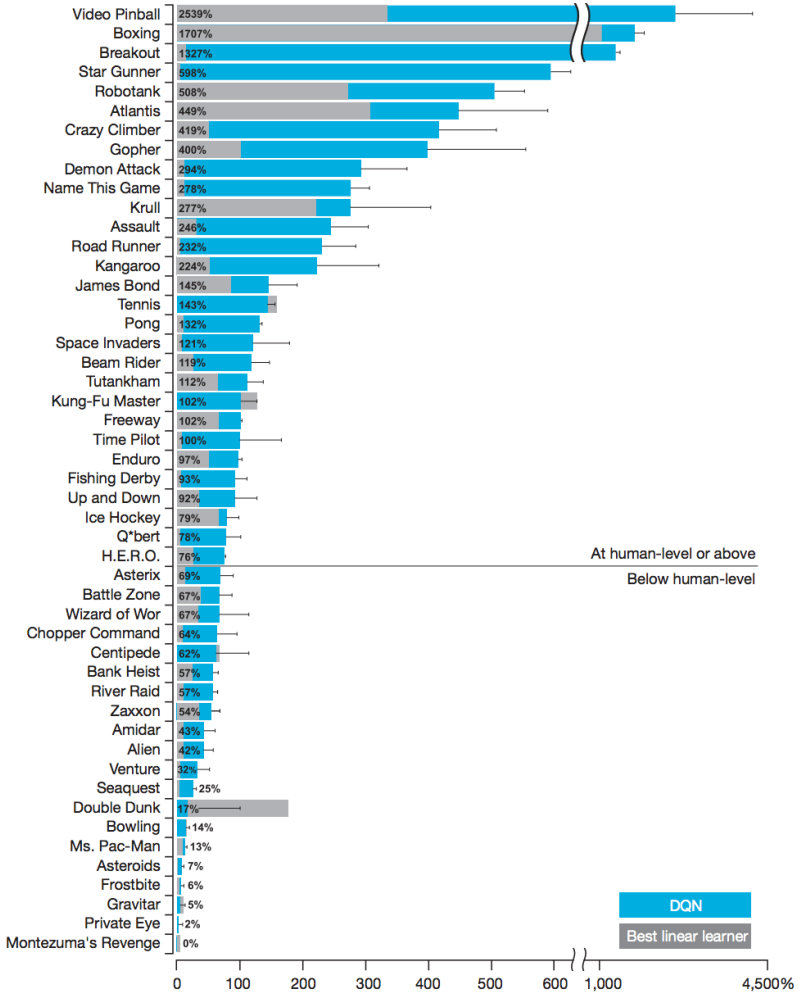
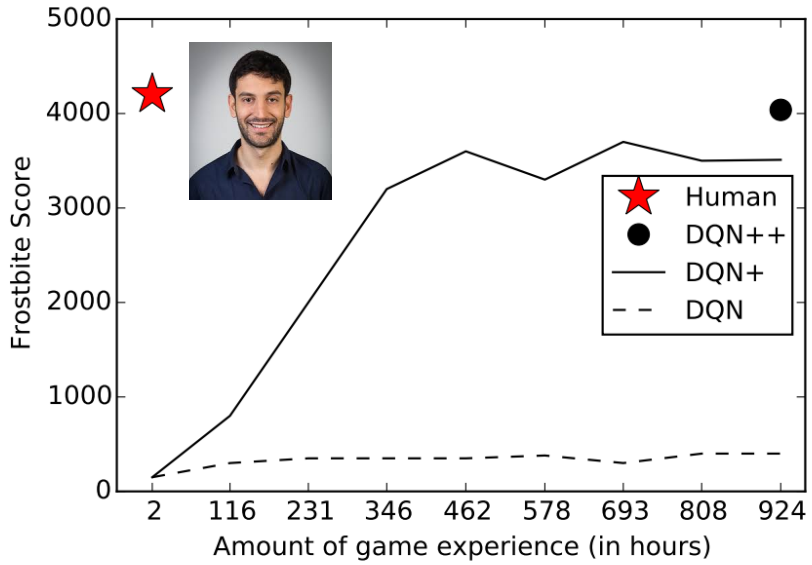
Reinforcement learning, fast and slow

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





Mnih *et al, Nature* (2015)



- Immediately exploit new information
- Integrate across observations
- Rich inference, leveraging past experience
- Strategic exploration

Lake *et al, BBS* (in press)

Hippocampal Contributions to Control: The Third Way

Máté Lengyel
Collegium Budapest Institute for Advanced Study
2 Szentháromság u, Budapest, H-1014, Hungary
and
Computational & Biological Learning Lab
Cambridge University Engineering Department
Trumpington Street, Cambridge CB2 1PZ, UK
lmate@gatsby.ucl.ac.uk

Peter Dayan
Gatsby Computational Neuroscience Unit, UCL
17 Queen Square, London WC1N 3AR, UK
dayan@gatsby.ucl.ac.uk

Abstract

Recent experimental studies have focused on the specialization of different neural structures for different types of instrumental behavior. Recent theoretical work has provided normative accounts for why there should be more than one control system, and how the output of different controllers can be integrated. Two particlar controllers have been identified, one associated with a forward model and the prefrontal cortex and a second associated with computationally simpler, habitual, actor-critic methods and part of the striatum. We argue here for the normative appropriateness of an additional, but so far marginalized control system, associated with episodic memory, and involving the hippocampus and medial temporal cortices. We analyze in depth a class of simple environments to show that episodic

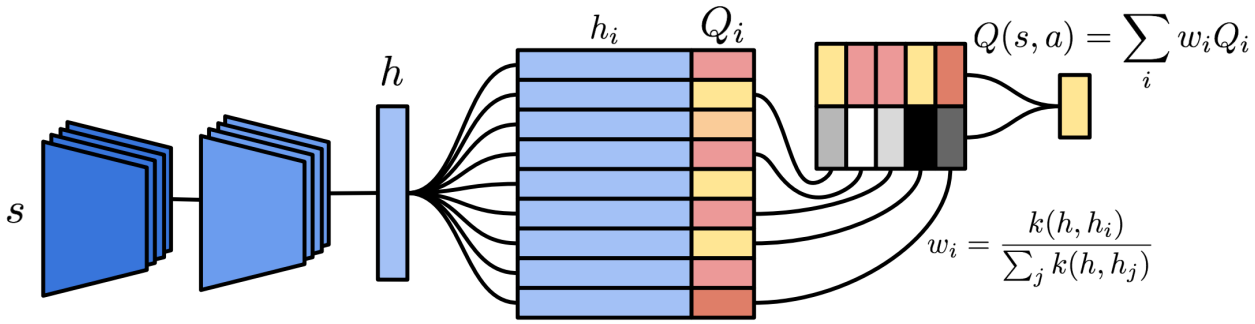
Learning to Learn Using Gradient Descent

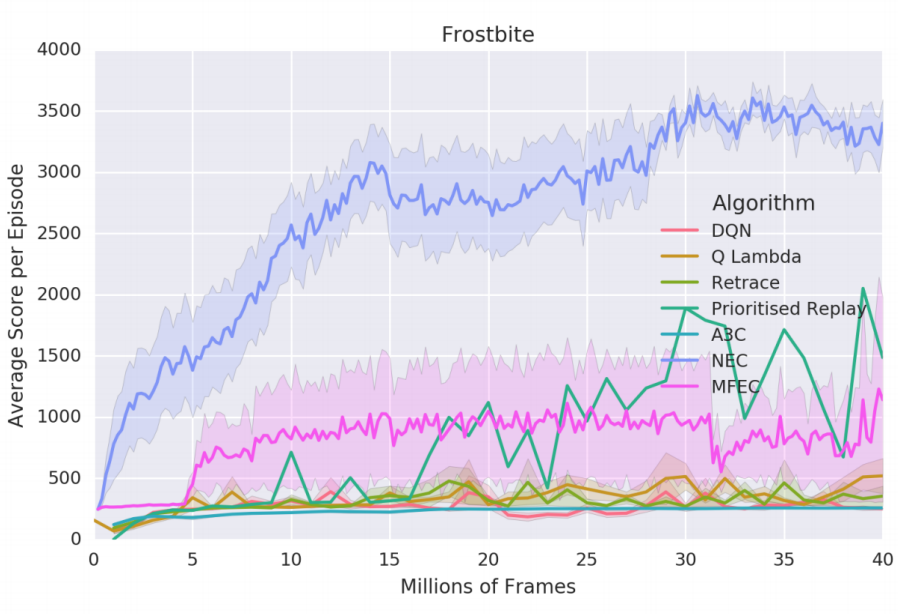
Sepp Hochreiter[1], A. Steven Younger[1], and Peter R. Conwell[2]

[1] Department of Computer Science
University of Colorado, Boulder, CO 80309–0430
[2] Physics Department
Westminster College, Salt Lake City, Utah

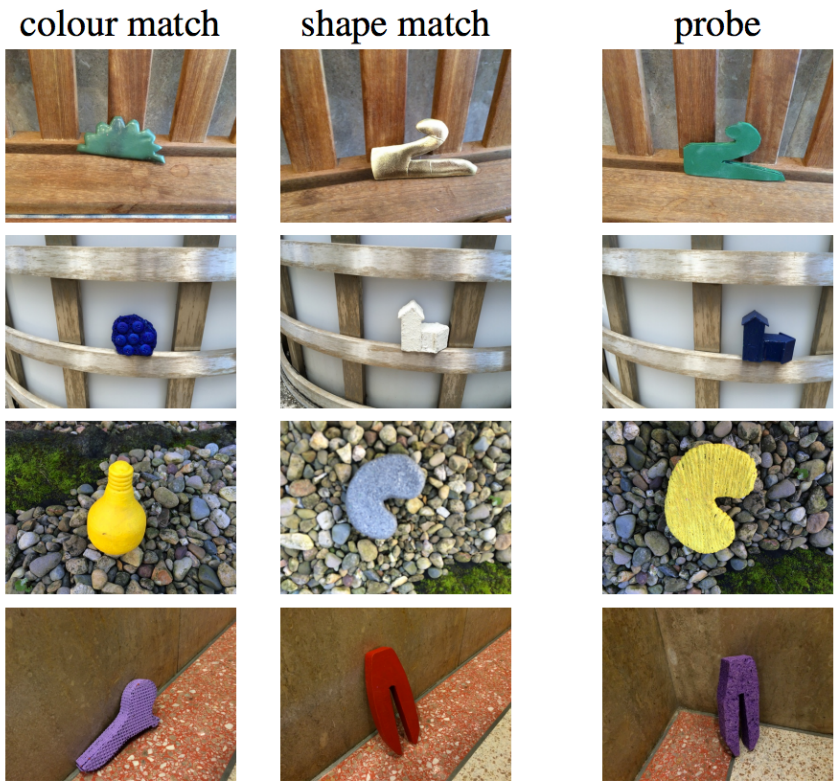
Abstract. This paper introduces the application of gradient descent methods to meta-learning. The concept of "meta-learning", i.e. of a system that improves or discovers a learning algorithm, has been of interest in machine learning for decades because of its appealing applications. Previous meta-learning approaches have been based on evolutionary methods and, therefore, have been restricted to small models with few free parameters. We make meta-learning in large systems feasible by using recurrent neural networks with their attendant learning routines as meta-learning systems. Our system derived complex well performing learning algorithms from scratch. In this paper we also show that our approach performs non-stationary time series prediction.

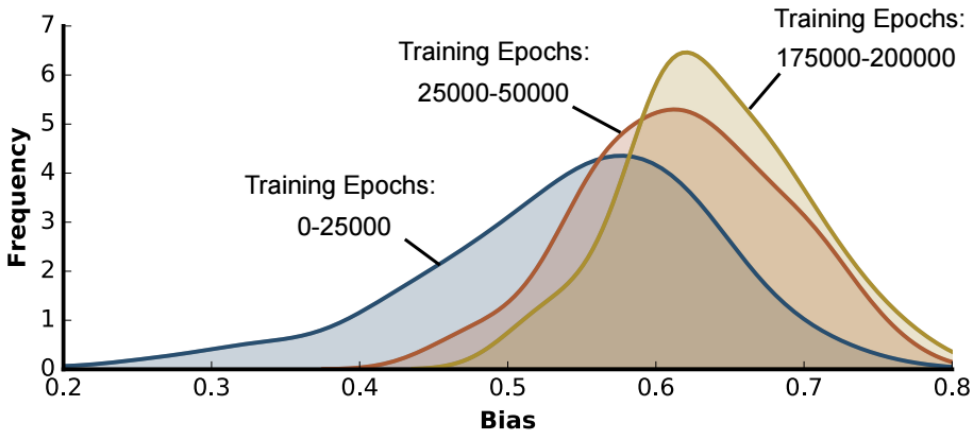
1 Introduction





Pritzl *et al, arXiv* (2017)





Ritter *et al.* (under review)

Hippocampal Contributions to Control: The Third Way

Máté Lengyel
Collegium Budapest Institute for Advanced Study
2 Szentháromság u, Budapest, H-1014, Hungary
and
Computational & Biological Learning Lab
Cambridge University Engineering Department
Trumpington Street, Cambridge CB2 1PZ, UK
lmate@gatsby.ucl.ac.uk

Peter Dayan
Gatsby Computational Neuroscience Unit, UCL
17 Queen Square, London WC1N 3AR, UK
dayan@gatsby.ucl.ac.uk

Abstract

Recent experimental studies have focused on the specialization of different neural structures for different types of instrumental behavior. Recent theoretical work has provided normative accounts for why there should be more than one control system, and how the output of different controllers can be integrated. Two particlar controllers have been identified, one associated with a forward model and the prefrontal cortex and a second associated with computationally simpler, habitual, actor-critic methods and part of the striatum. We argue here for the normative appropriateness of an additional, but so far marginalized control system, associated with episodic memory, and involving the hippocampus and medial temporal cortices. We analyze in depth a class of simple environments to show that episodic

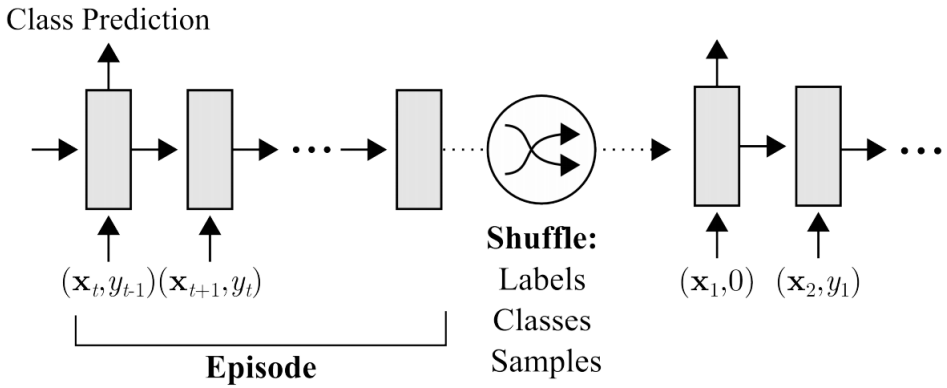
Learning to Learn Using Gradient Descent

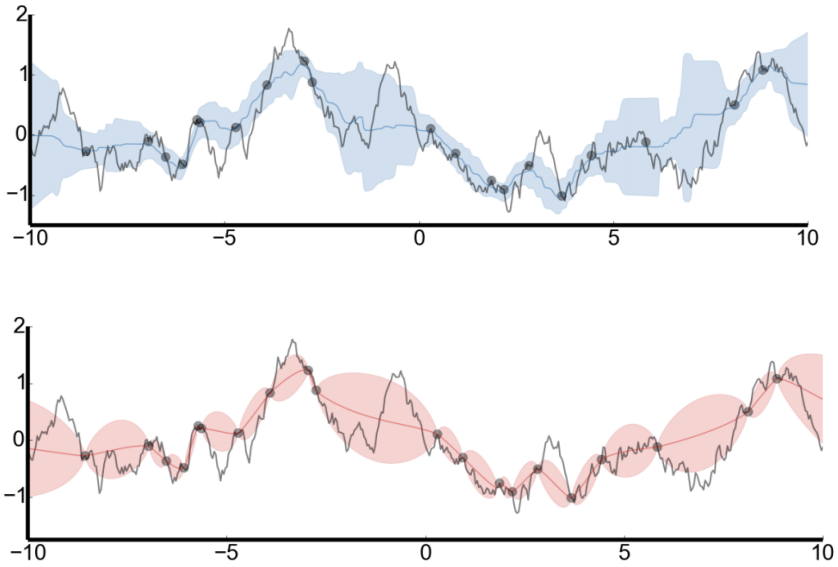
Sepp Hochreiter[1], A. Steven Younger[1], and Peter R. Conwell[2]

[1] Department of Computer Science
University of Colorado, Boulder, CO 80309–0430
[2] Physics Department
Westminster College, Salt Lake City, Utah

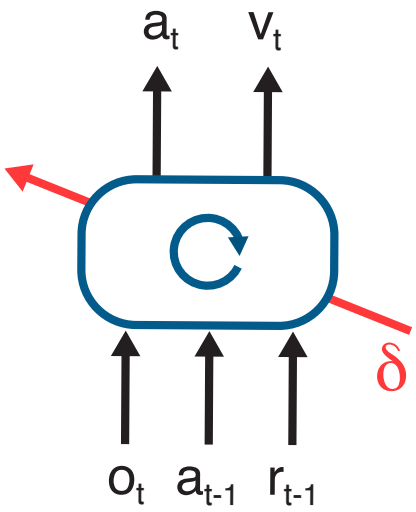
Abstract. This paper introduces the application of gradient descent methods to meta-learning. The concept of "meta-learning", i.e. of a system that improves or discovers a learning algorithm, has been of interest in machine learning for decades because of its appealing applications. Previous meta-learning approaches have been based on evolutionary methods and, therefore, have been restricted to small models with few free parameters. We make meta-learning in large systems feasible by using recurrent neural networks with their attendant learning routines as meta-learning systems. Our system derived complex well performing learning algorithms from scratch. In this paper we also show that our approach performs non-stationary time series prediction.

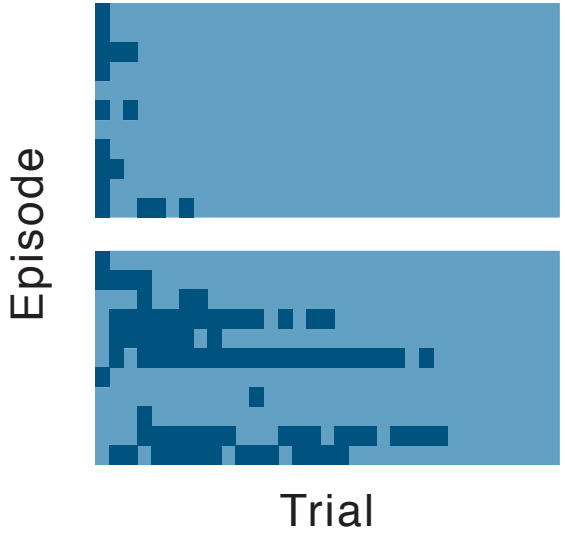
1 Introduction

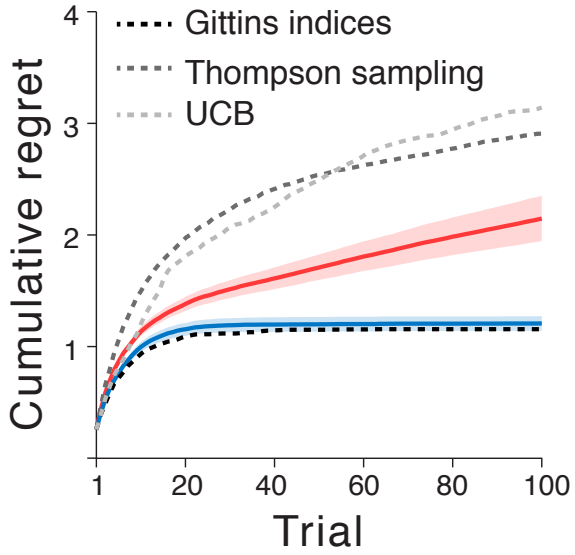




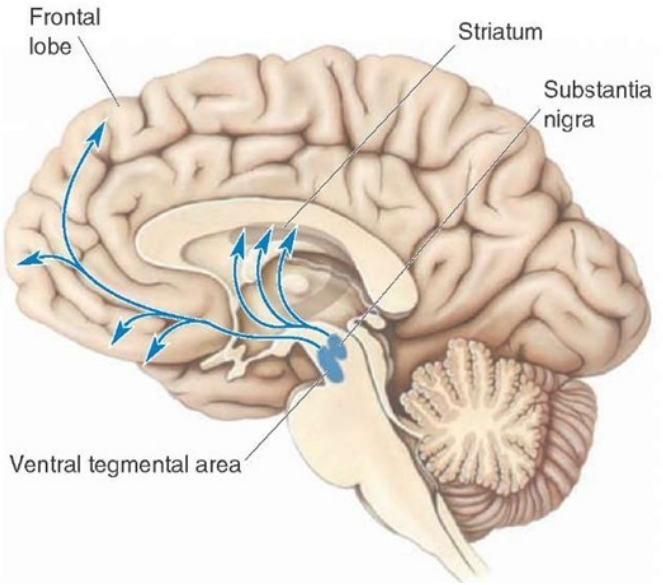
Santoro *et al.*, *ICML* (2016)

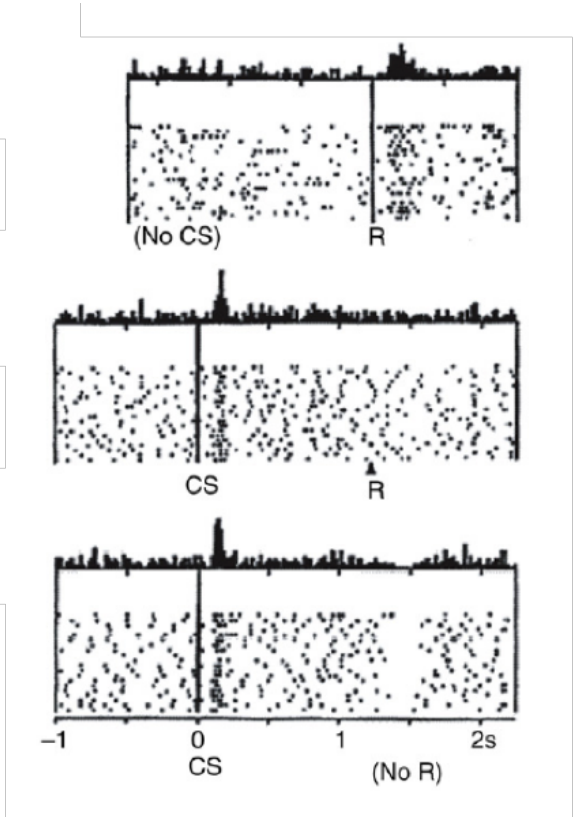


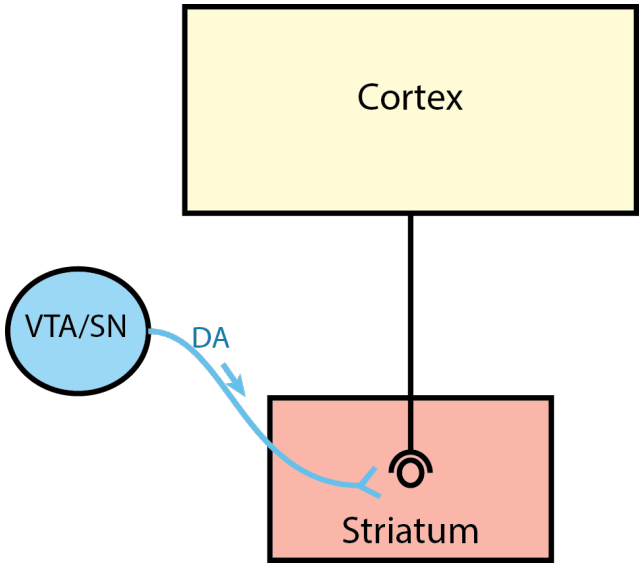




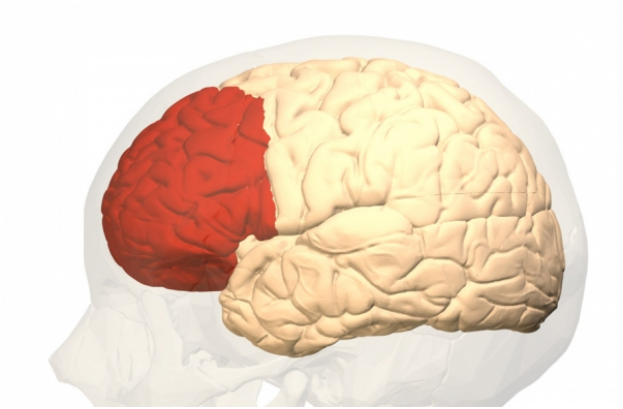
Wang *et al.,* arXiv (2016)
Duan et al., arXiv (2016)

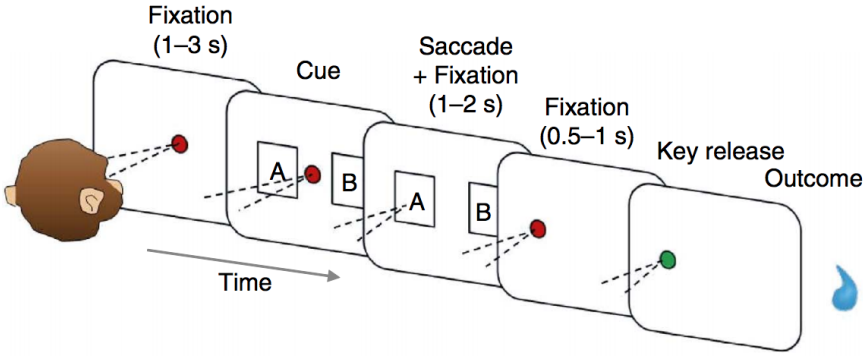


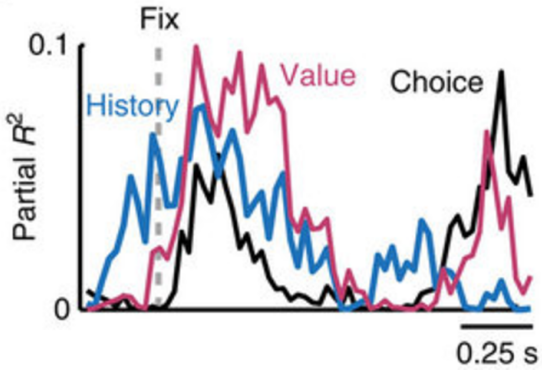




Schultz et al, Science (1997)

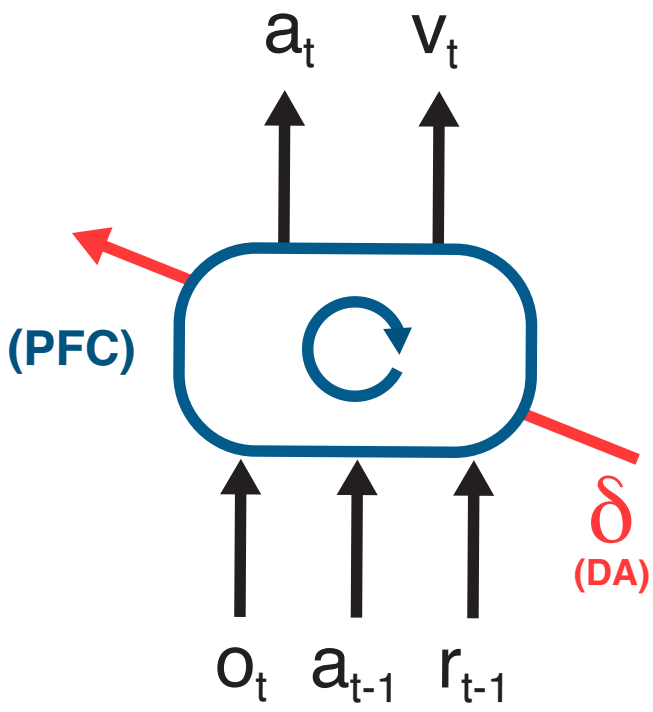




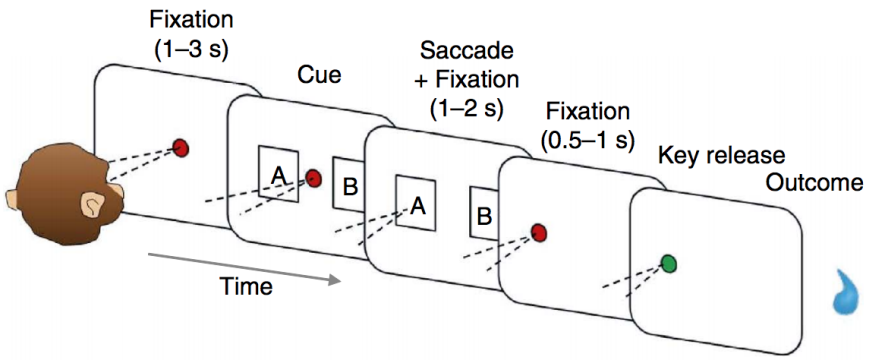


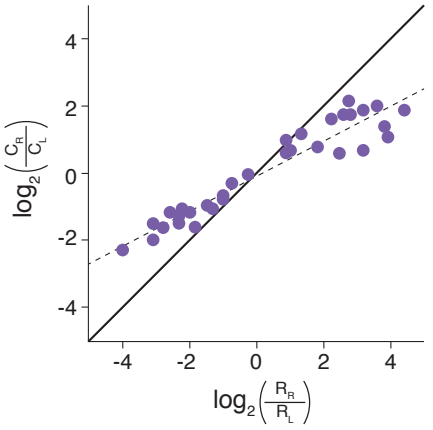
Tsutsui et al., Nat. Comm., 2016

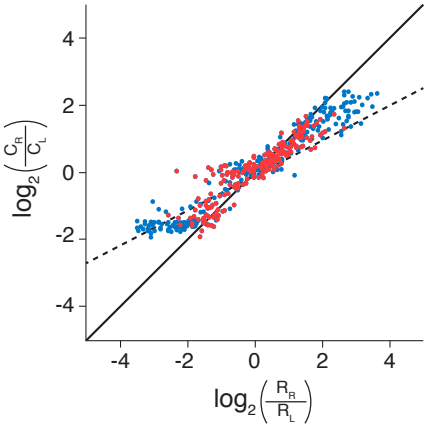
See also:
Barraclough et al., Nat. Neuro. 2004
Seo & Lee, J. Neurosci 2007
Shima & Tanji, Science 1998
Matsumoto et al., Nat. Neuro. 2007

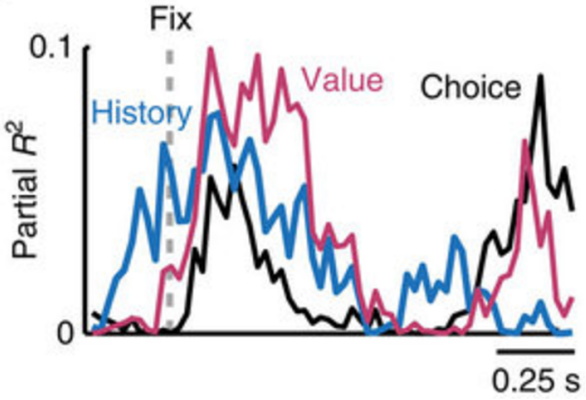


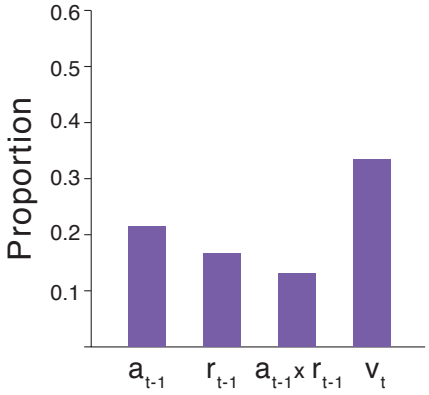
Wang *et al.*, arXiv (2016)

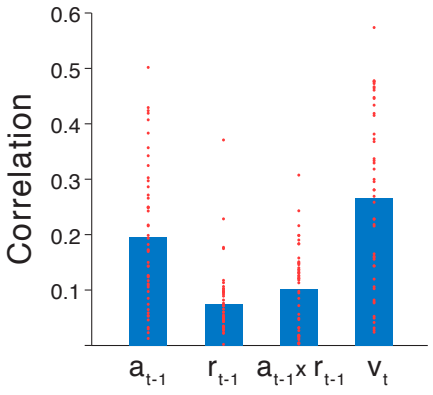






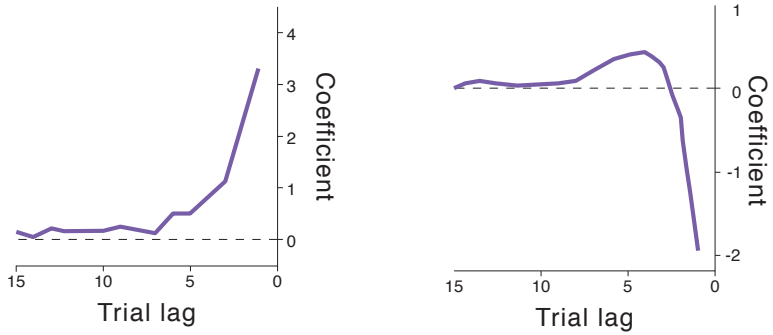






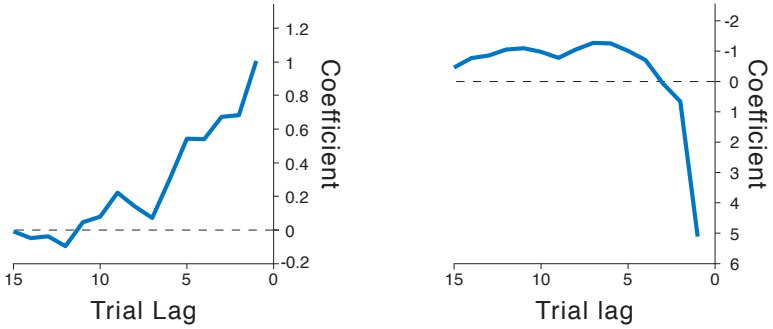
Wang *et al.* (under review)

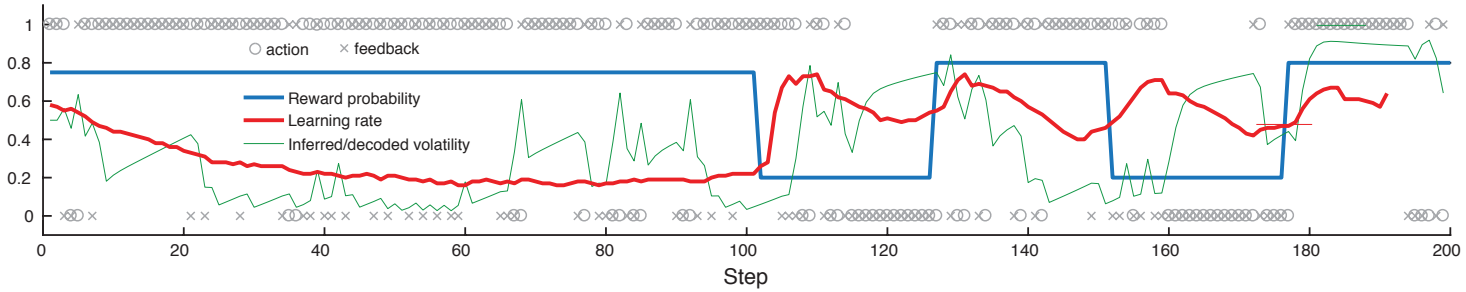
Effects of past outcome and choice on next choice

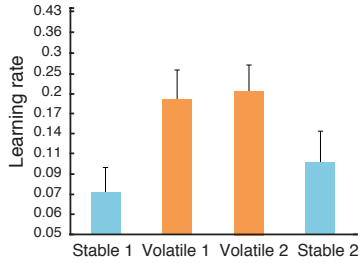


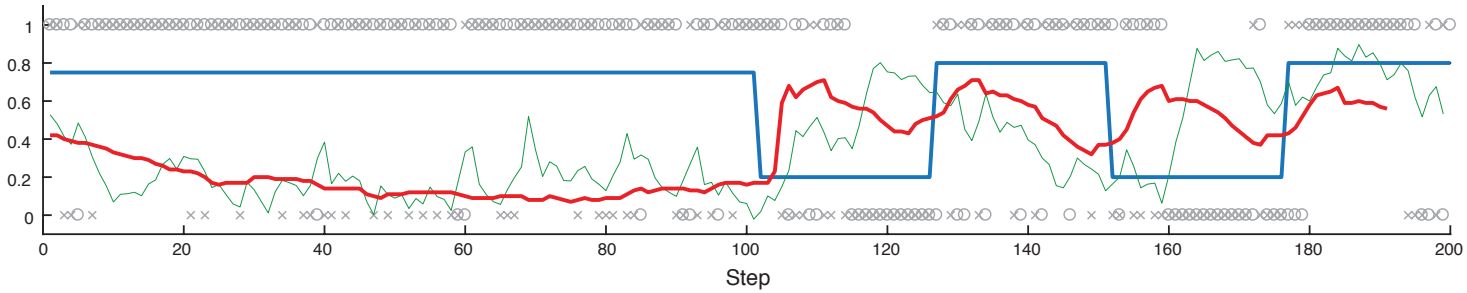
Lau and Glimcher, J. Exp. Analysis Behavior 2005

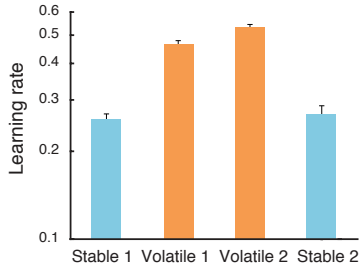
Meta-RL



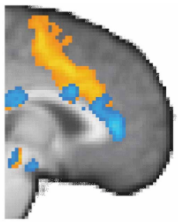


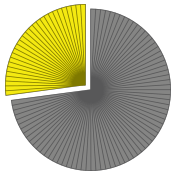






Units coding for volatility

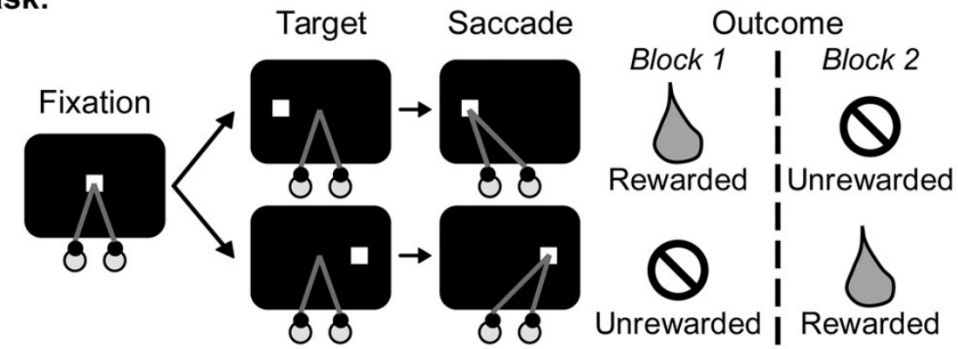




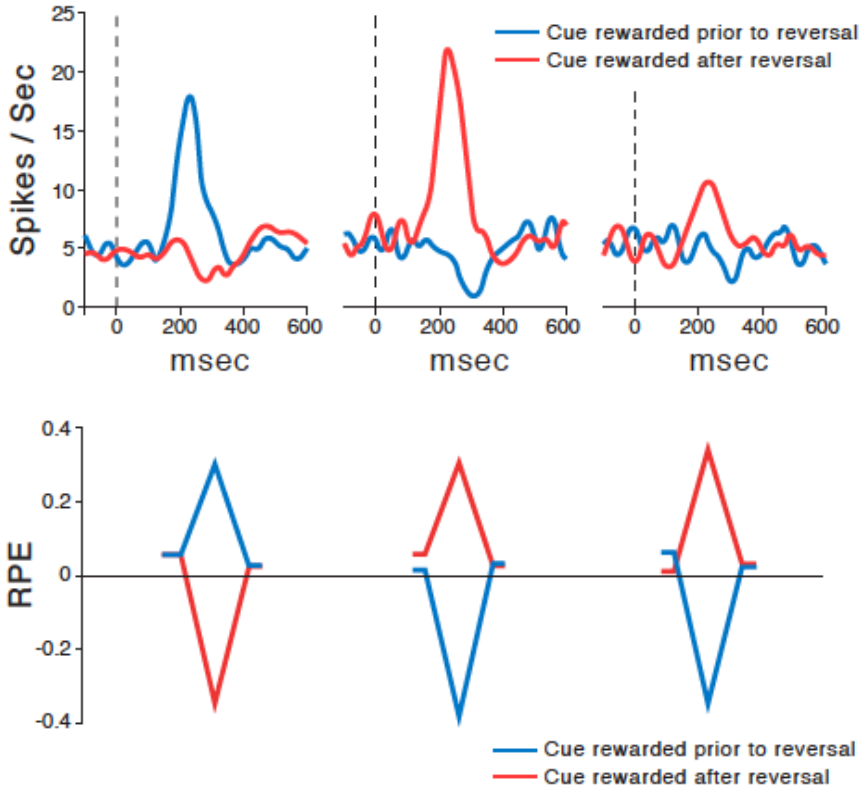
Behrens et al., Nature Neuroscience, 2007

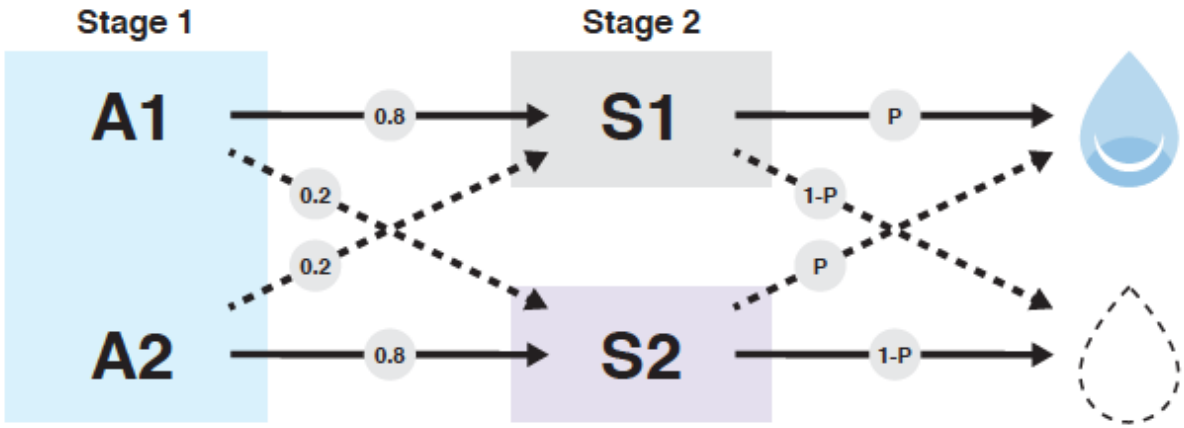
Wang *et al.* (under review)

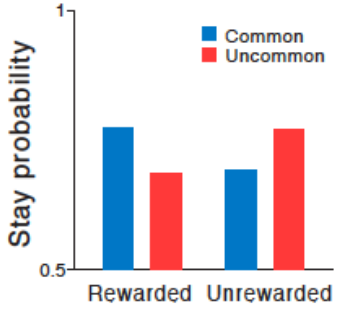
Task:

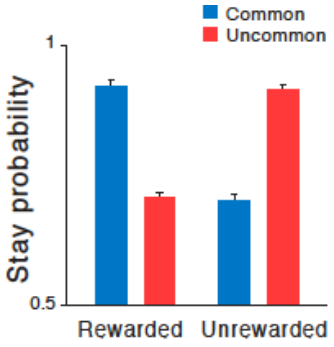


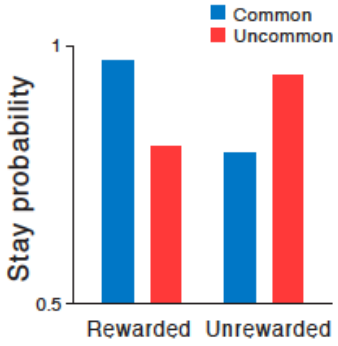
Bromberg-Martin et al, J Neurophys, 2010



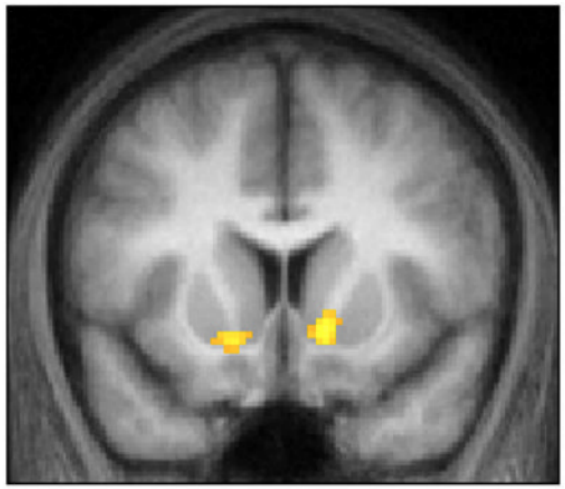


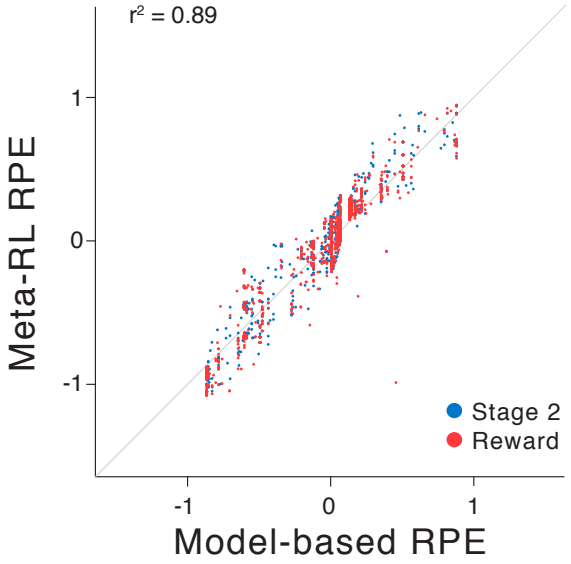




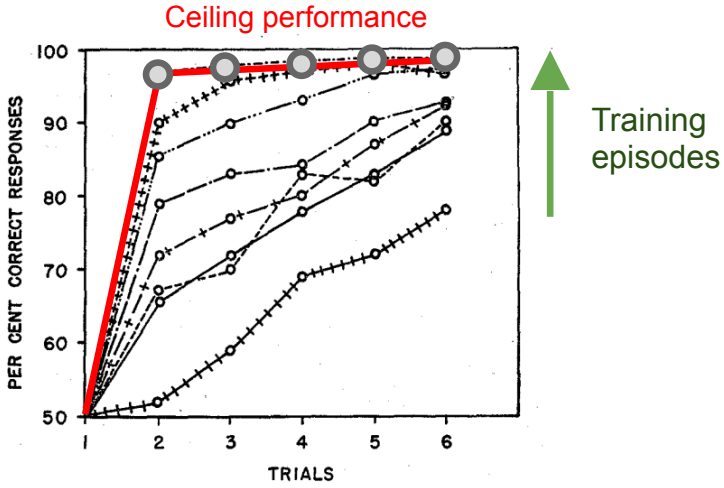


Miller, Botvinick & Brody (under review); Daw et al., Neuron, 2011

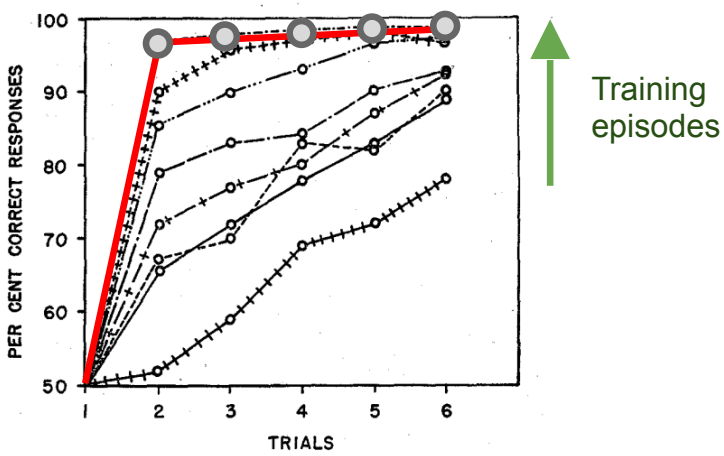




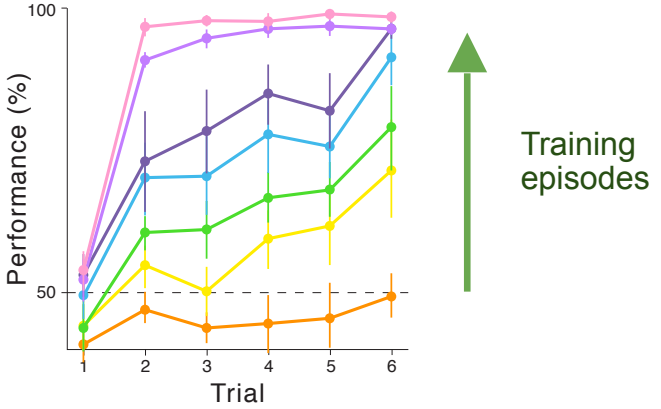
Daw et al., Neuron, 2011



Harlow, Psychological Review, 1949



Meta-RL

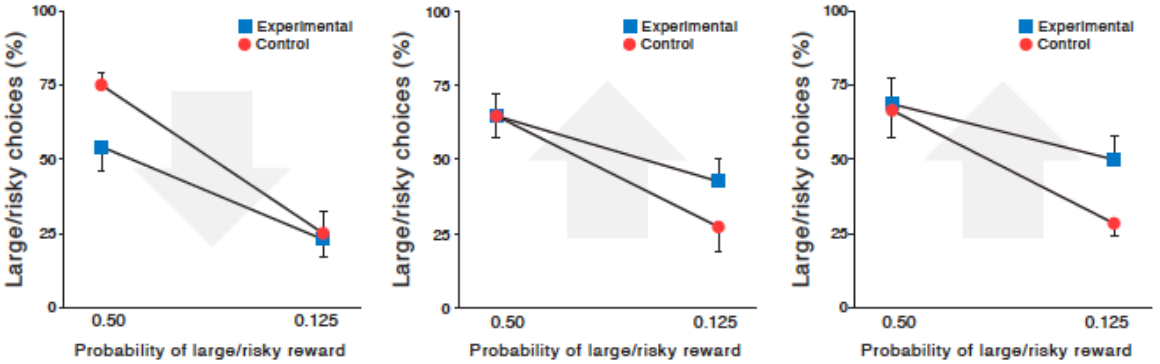


Harlow, Psychological Review, 1949

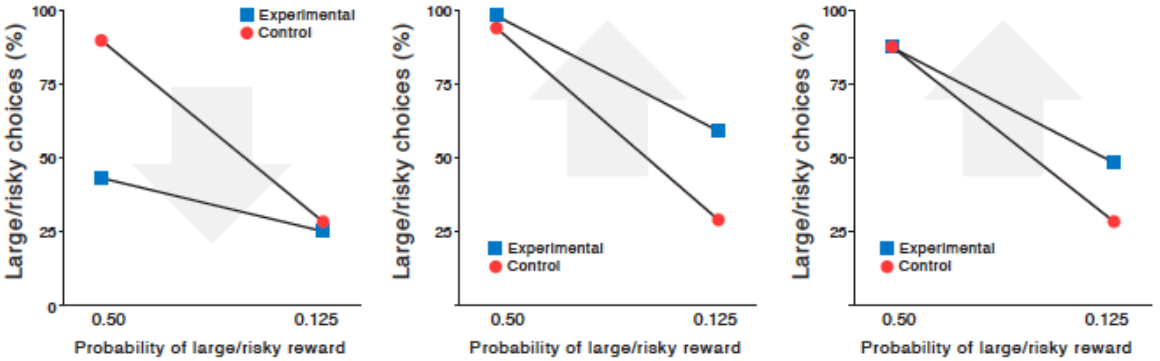
DA blocked upon food reward from large/risky option

DA blocked upon food reward from small/certain option

DA triggered upon food omission from large/risky option



E



Stopper et al., Neuron, 2014

Thank you!

Jane Wang
Zeb Kurth-Nelson
Dharshan Kumaran
Chris Summerfield
Hubert Soyer
Joel Leibo
Sam Ritter
Adam Santoro
Tim Lillicrap
David Barrett
Dhruva Tirumala
Remi Munos
Charles Blundell
Demis Hassabis