Tools that learn

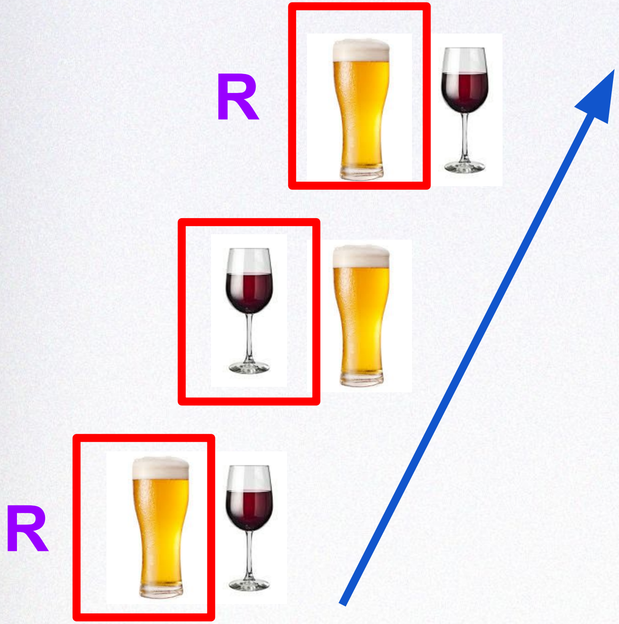
Nando de Freitas and many DeepMind colleagues

DeepMind

Learning **slow** to learn **fast**

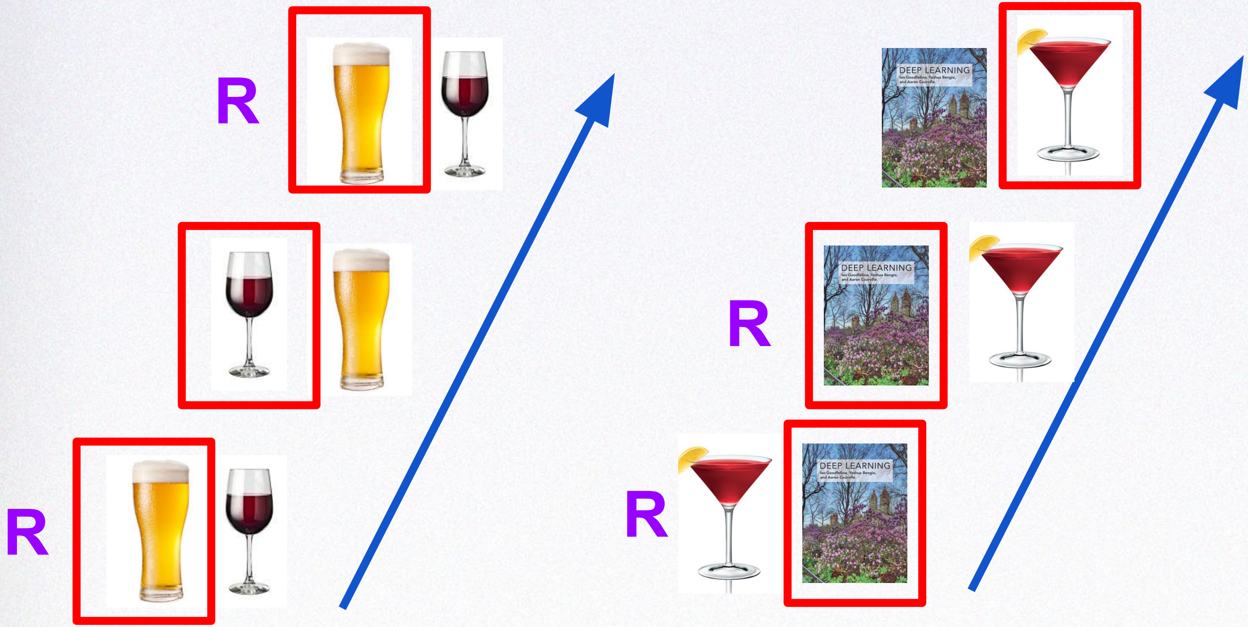
- Infants are endowed with systems of core knowledge for reasoning about objects, actions, number, space, and social interactions [eg E. Spelke].
- The slow learning process of evolution led to the emergence of components that enable fast and varied forms of learning.

Harlow showed a monkey 2 visually contrasting objects. One covering food, the other nothing. The monkey chose between the 2. The process continued for a set number of trials using the same 2 objects, then again with 2 different objects.



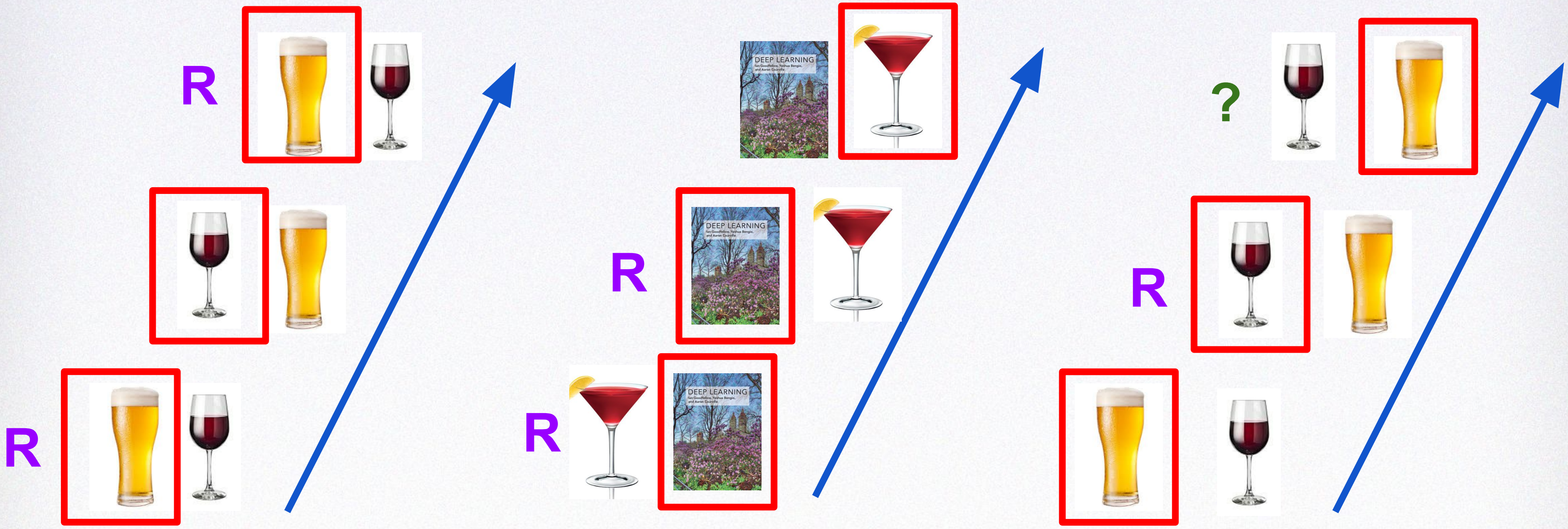
Harlow (1949), Jane Wang et al (2016)

Harlow showed a monkey 2 visually contrasting objects. One covering food, the other nothing. The monkey chose between the 2. The process continued for a set number of trials using the same 2 objects, then again with 2 different objects.



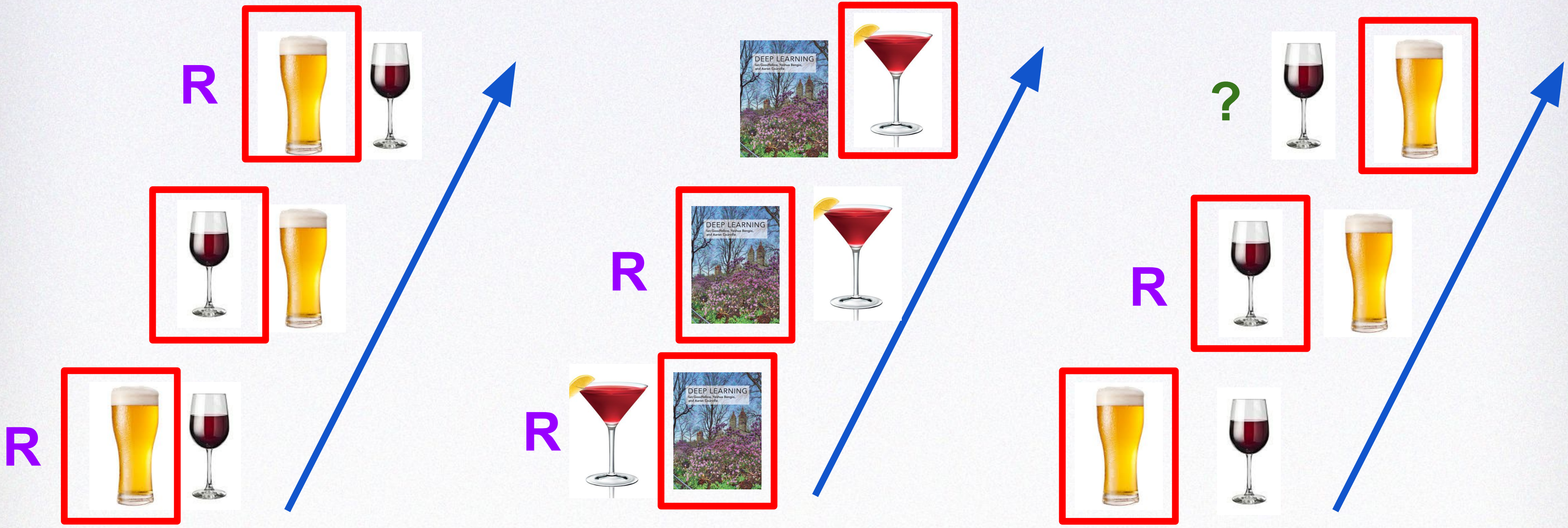
Harlow (1949), Jane Wang et al (2016)

Harlow showed a monkey 2 visually contrasting objects. One covering food, the other nothing. The monkey chose between the 2. The process continued for a set number of trials using the same 2 objects, then again with 2 different objects.



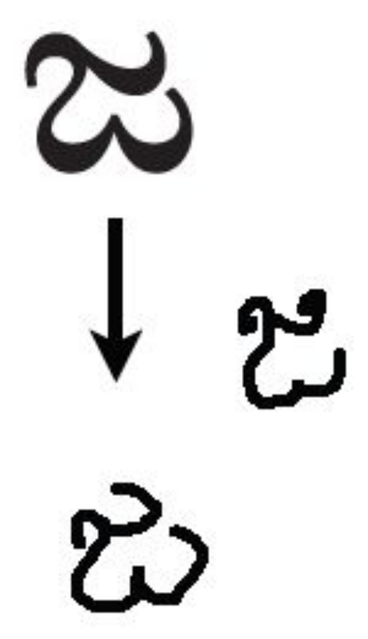
Harlow (1949), Jane Wang et al (2016)

Harlow showed a monkey 2 visually contrasting objects. One covering food, the other nothing. The monkey chose between the 2. The process continued for a set number of trials using the same 2 objects, then again with 2 different objects.



Eventually, when 2 new objects were presented, the monkey’s first choice between them was arbitrary. But after observing the outcome of the first choice, the monkey would subsequently always choose the right one. Harlow (1949), Jane Wang et al (2016)

Learning to learn is intimately related to few shot learning



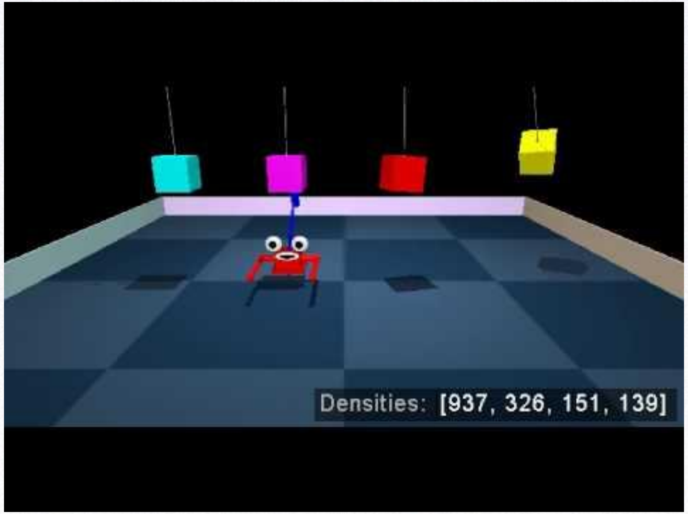
- **Challenge**: how can a neural net learn from few examples?
- **Answer**: Learn a model that expects a few data at test time, and knows how to capitalize on this data.

Brenden Lake et al (2016) Adam Santoro et al (2016)
… Hugo Larochelle, Chelsea Finn, and many others

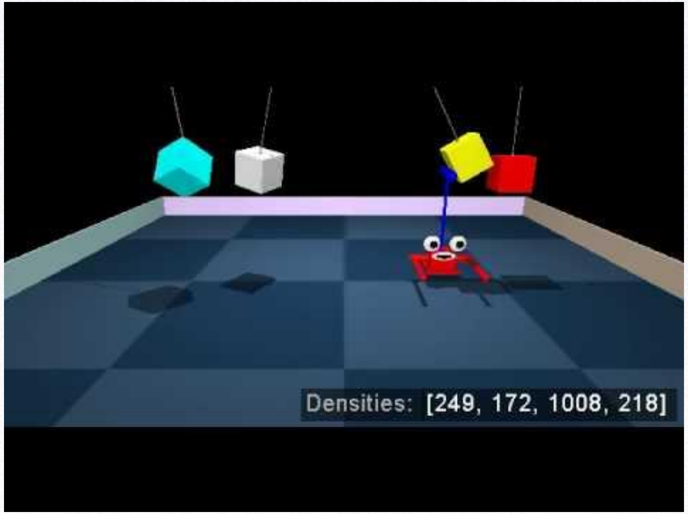
Learn to experiment

Agent learns to solve bandit problems with meta RL

Before learning



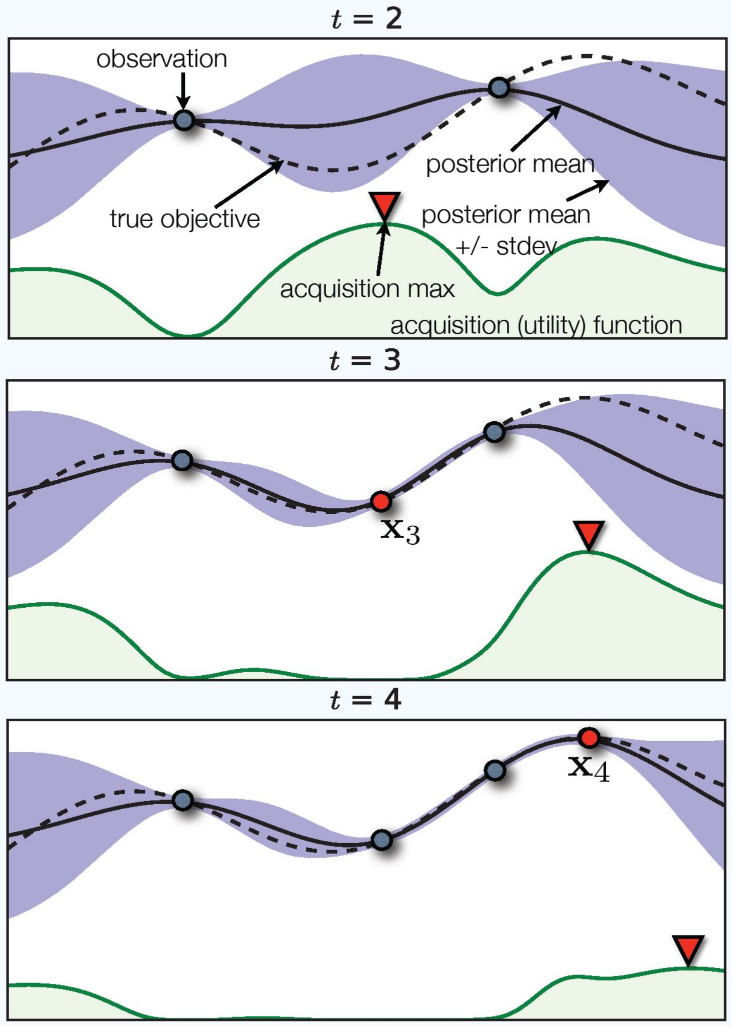
After learning

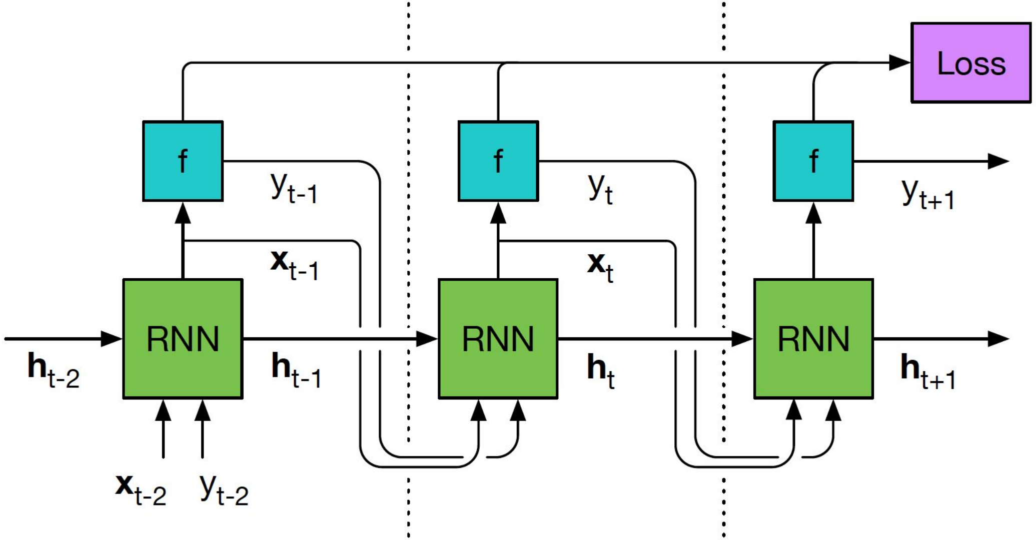


Misha Denil, Pulkit Agrawal, Tejas Kulkarni, Tom Erez, Peter Battaglia, NdF (2017)

Learn to optimize

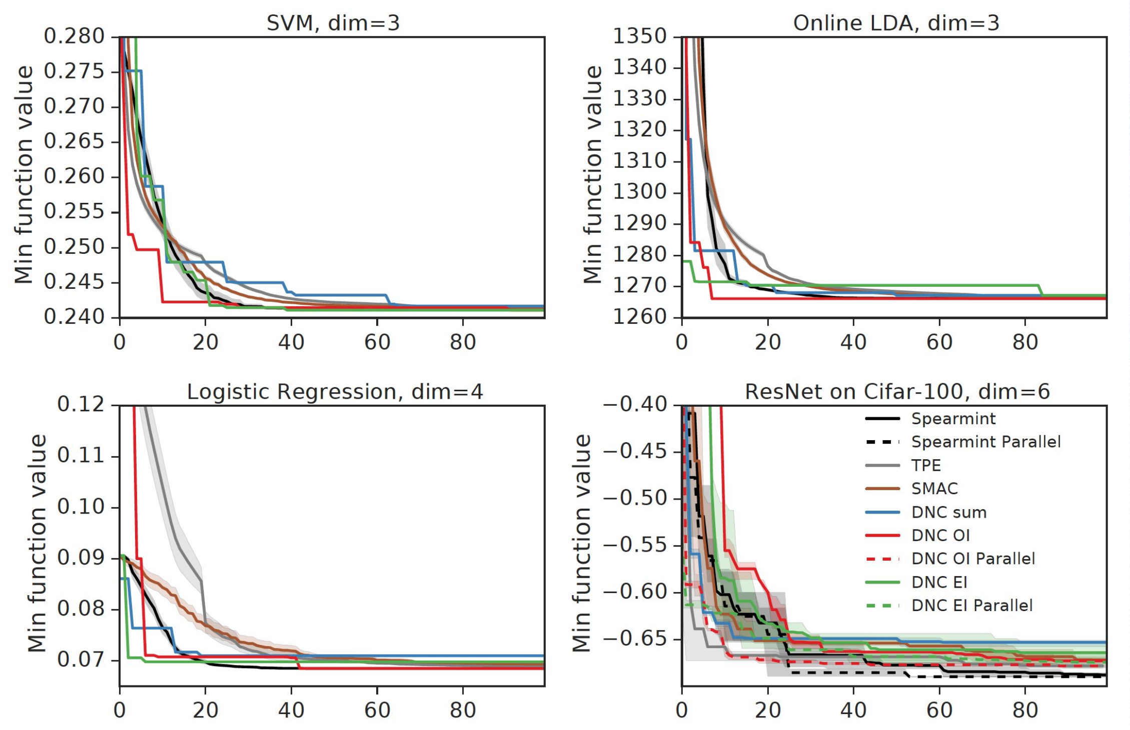
Neural Bayesian optimization

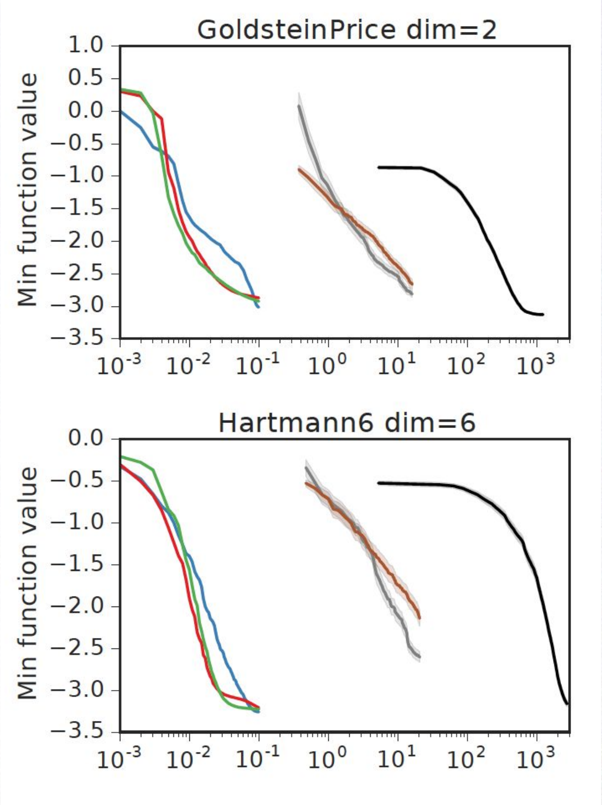




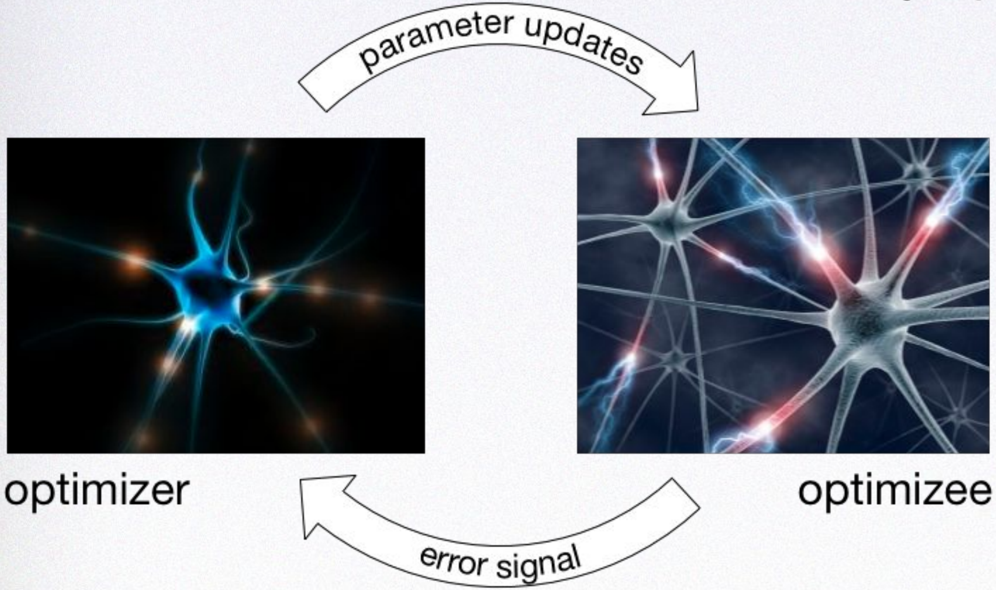
Yutian Chen, Matthew Hoffman, Sergio Gomez,
Misha Denil, Timothy Lillicrap, Matt Botvinick, NdF (2017)

Transfer to hyper-parameter optimization in ML





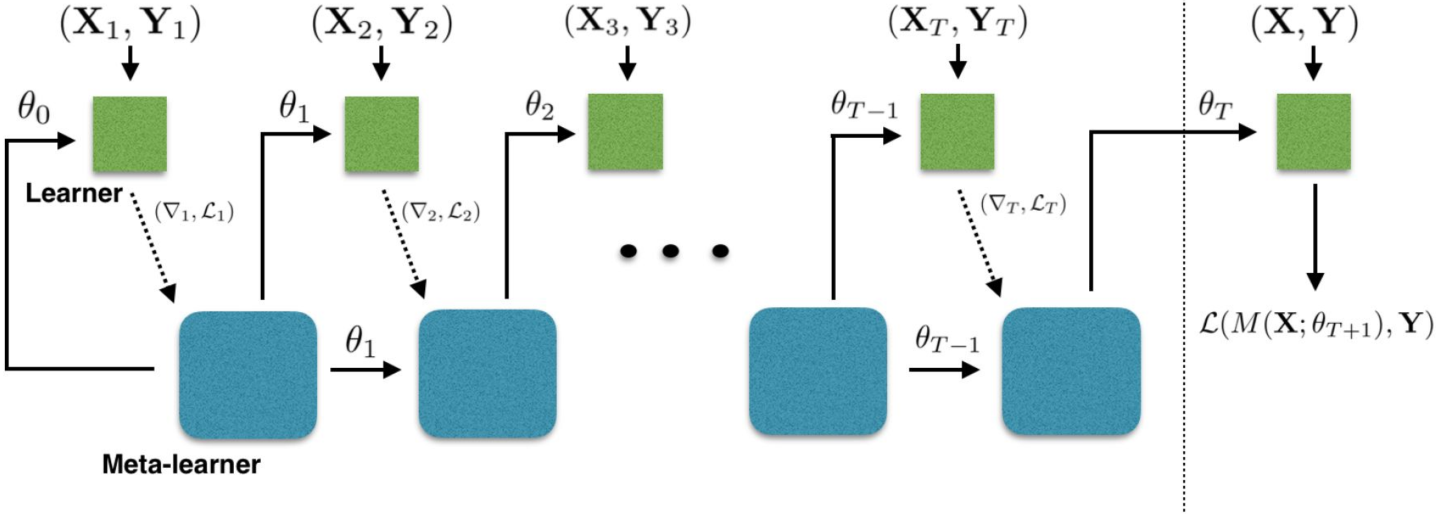
Learning to learn by gradient descent by gradient descent



$$\theta_{t+1} = \theta_t + g_t(\nabla f(\theta_t), \phi)$$

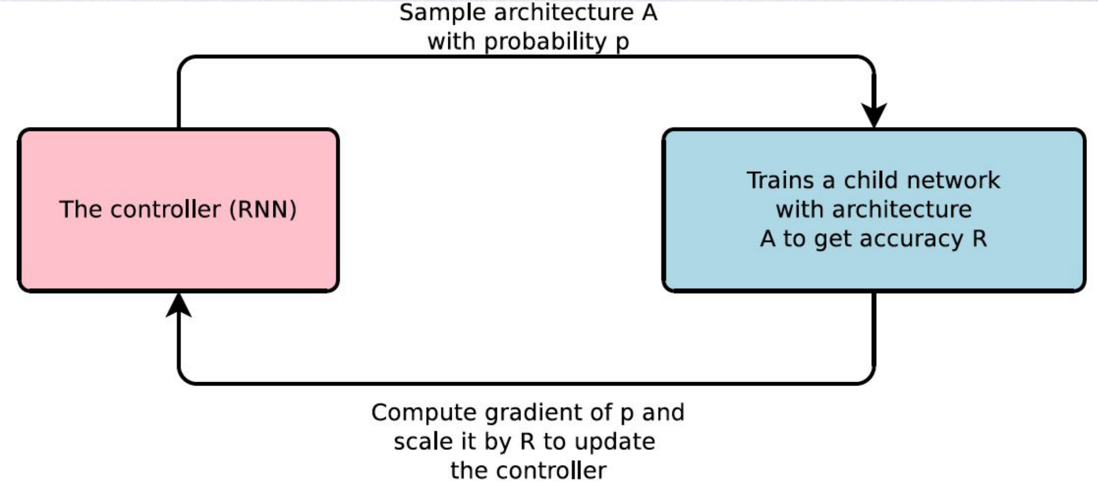
Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, NdF (2016)

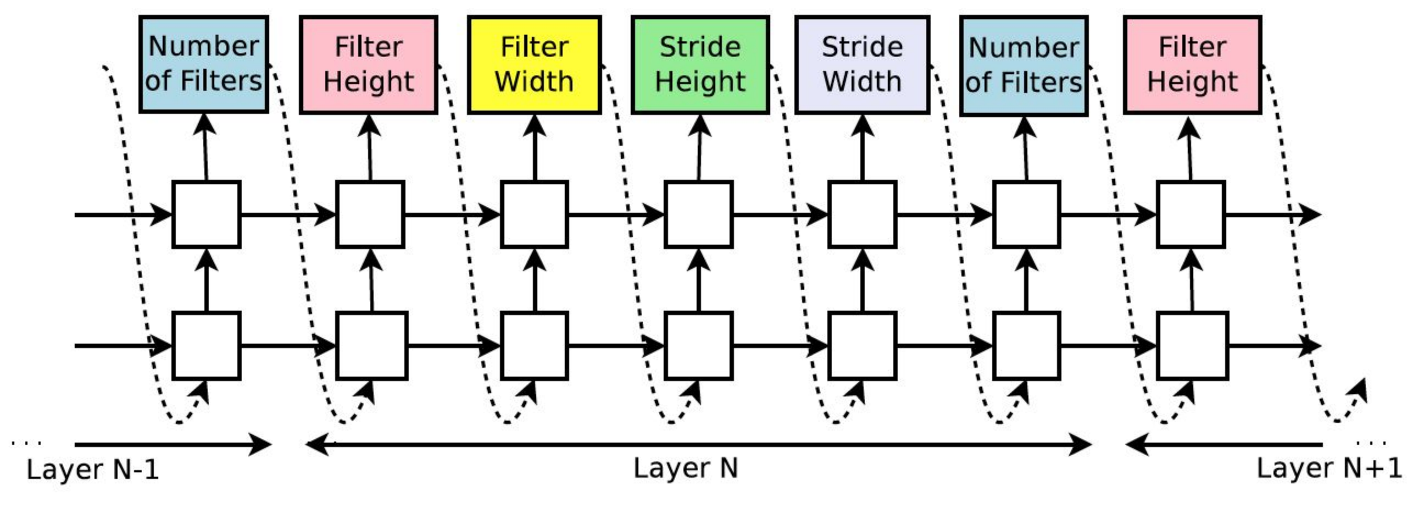
Few-shot learning to learn



Sachin Ravi, Hugo Larochelle (2017)

Architecture search

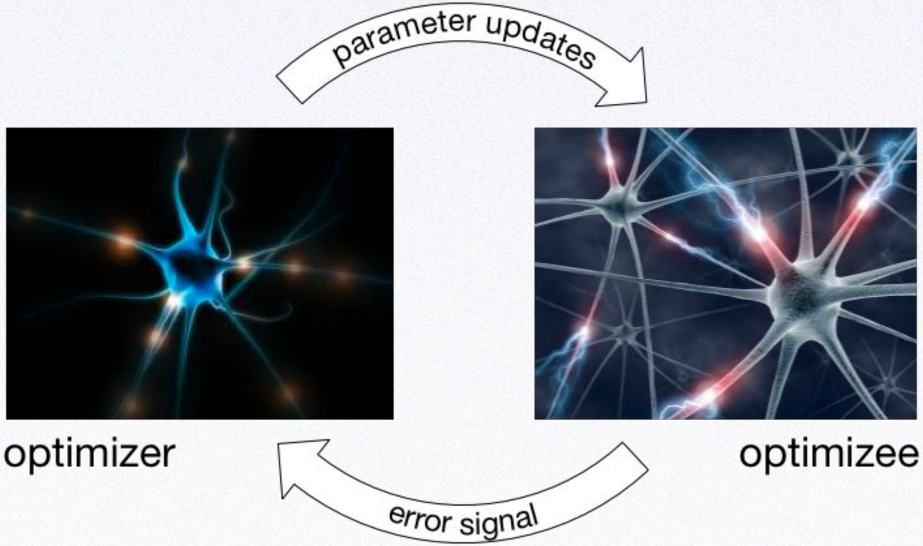




Barret Zoph and Quoc Le (2017)

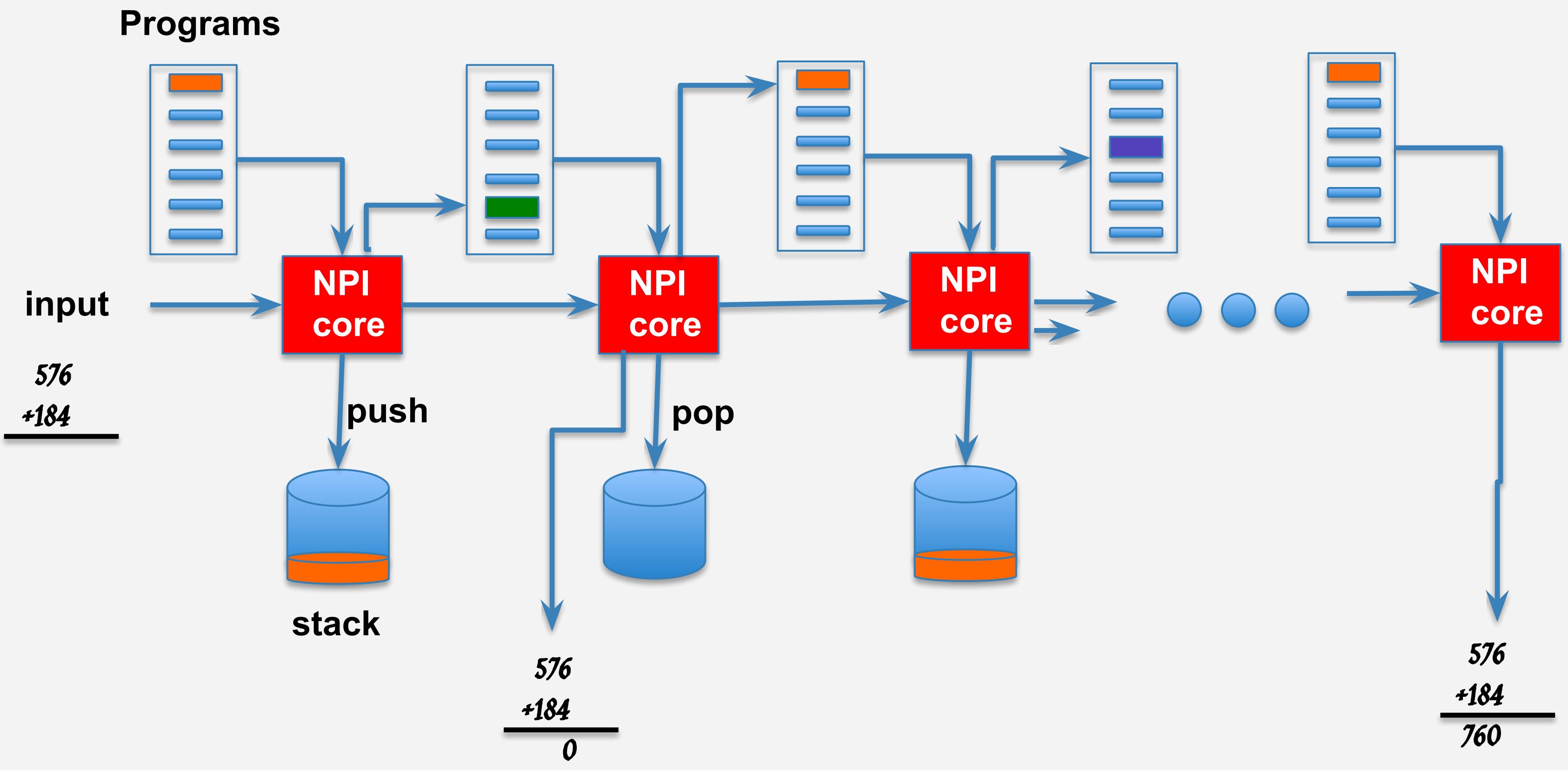
Learn to program

Networks programming other networks



McClelland, Rumelhart and Hinton (1987)

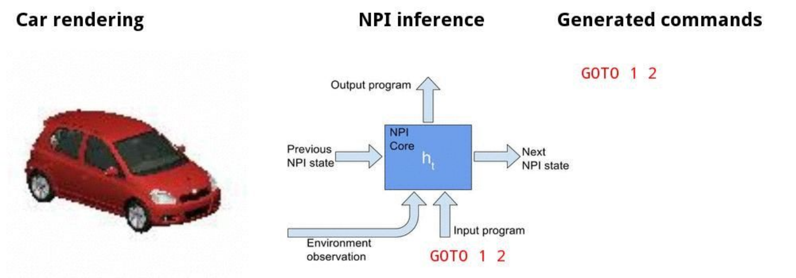
NPI – a net with recursion that learns a **finite** set of programs

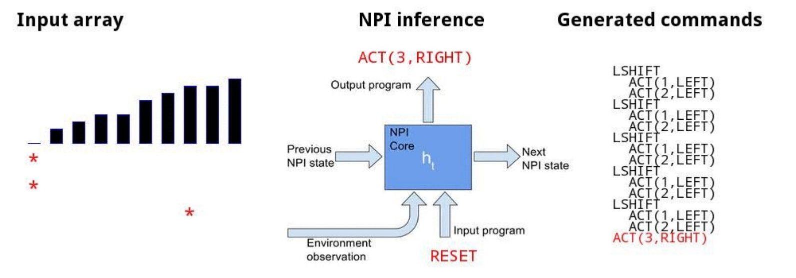


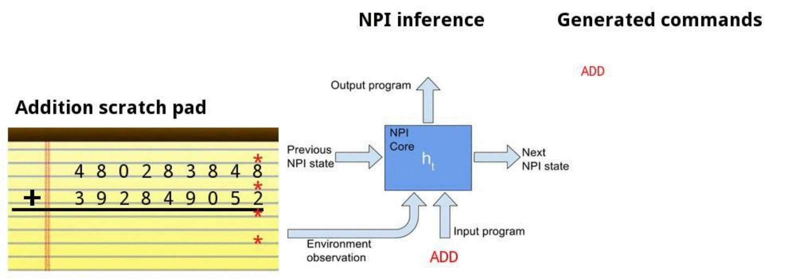
Google DeepMind

Reed and NdF [2016]

Multi-task: Same network and same core parameters







Program	Descriptions	Calls
ADD	Perform multi-digit addition	ADD1, LSHIFT
ADD1	Perform single-digit addition	ACT, CARRY
CARRY	Mark a 1 in the carry row one unit left	ACT
LSHIFT	Shift a specified pointer one step left	ACT
RSHIFT	Shift a specified pointer one step right	ACT
ACT	Move a pointer or write to the scratch pad	-
BUBBLESORT	Perform bubble sort (ascending order)	BUBBLE, RESET
BUBBLE	Perform one sweep of pointers left to right	ACT, BSTEP
RESET	Move both pointers all the way left	LSHIFT
BSTEP	Conditionally swap and advance pointers	COMPSWAP, RSHIFT
COMPSWAP	Conditionally swap two elements	ACT
LSHIFT	Shift a specified pointer one step left	ACT
RSHIFT	Shift a specified pointer one step right	ACT
ACT	Swap two values at pointer locations or move a pointer	-
GOTO	Change 3D car pose to match the target	HGOTO, VGOTO
HGOTO	Move horizontally to the target angle	LGOTO, RGOTO
LGOTO	Move left to match the target angle	ACT
RGOTO	Move right to match the target angle	ACT
VGOTO	Move vertically to the target elevation	UGOTO, DGOTO
UGOTO	Move up to match the target elevation	ACT
DGOTO	Move down to match the target elevation	ACT
ACT	Move camera 15° up, down, left or right	-
RJMP	Move all pointers to the rightmost posiiton	RSHIFT
MAX	Find maximum element of an array	BUBBLESORT,RJMP

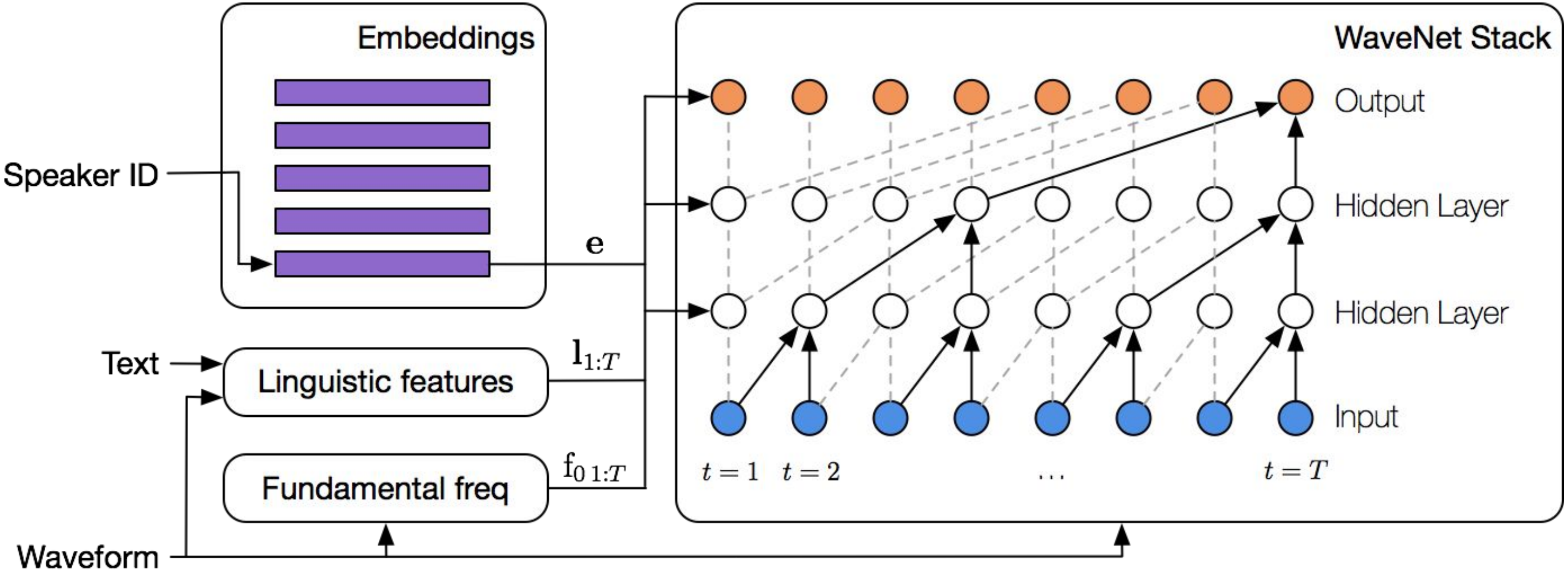
Meta-learning: Learning new programs with a fixed NPI core

- Maximum-finding in an array. Simple solution: Call BUBBLESORT and then take the rightmost element.
- Learn the new program by backpropagation with the NPI core and all other parameters fixed.

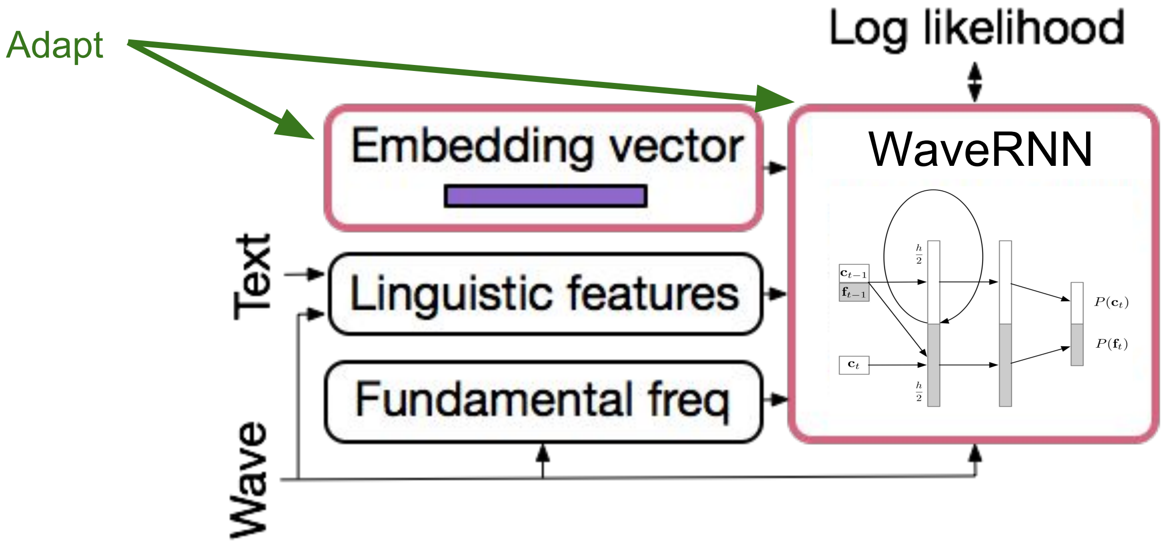
Task	Single	Multi	+ Max
Addition	100.0	97.0	97.0
Sorting	100.0	100.0	100.0
Canon. seen car	89.5	91.4	91.4
Canon. unseen	88.7	89.9	89.9
Maximum	-	-	100.0

Learn to imitate

Few-shot text to speech

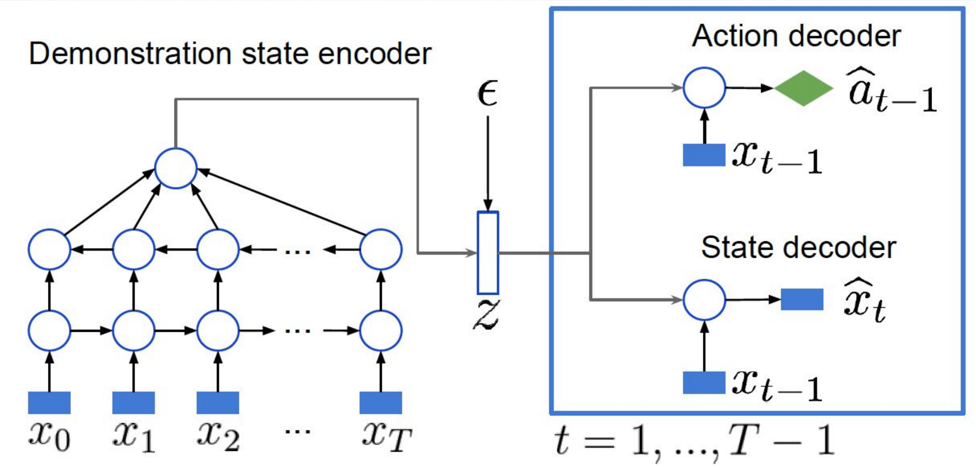


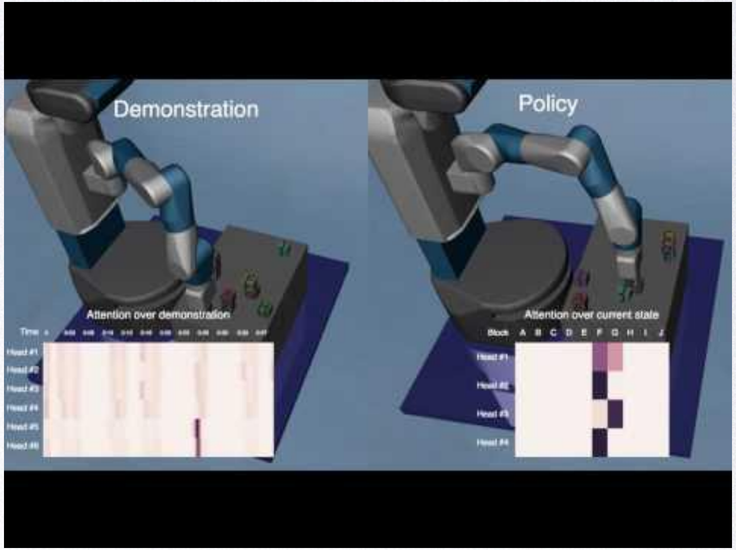
Same Adaptation Applies to **WaveRNN**

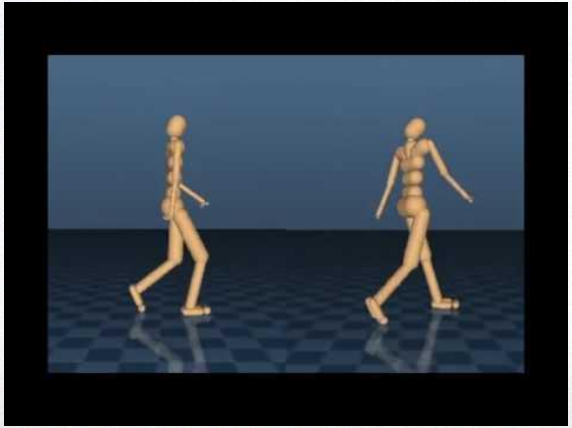


- Few-shot WaveNet and WaveRNN achieve the same sample quality (with 5 minutes) as the model trained from scratch with 4 hours of data.

One-shot imitation learning







Ziyu Wang, Josh Merel, Scott Reed, Greg Wayne, NdF, Nicolas Heess (2017)

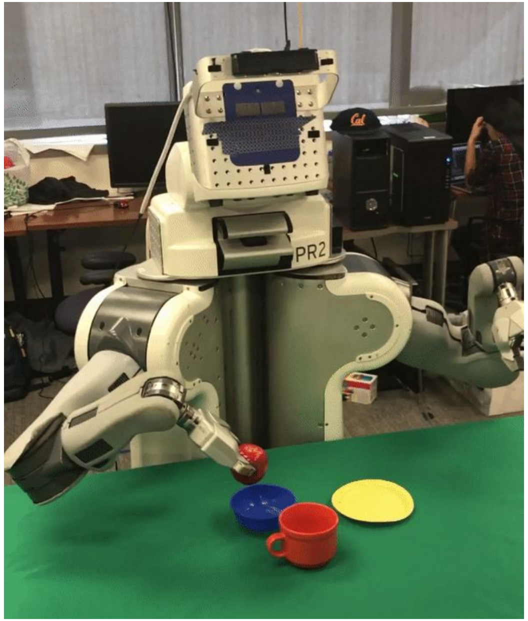
Yan Duan, Marcin Andrychowicz, Bradly Stadie, Jonathan Ho, Jonas Schneider, Ilya Sutskever, Pieter Abbeel, Wojciech Zaremba (2017)

One-Shot Imitation Learning

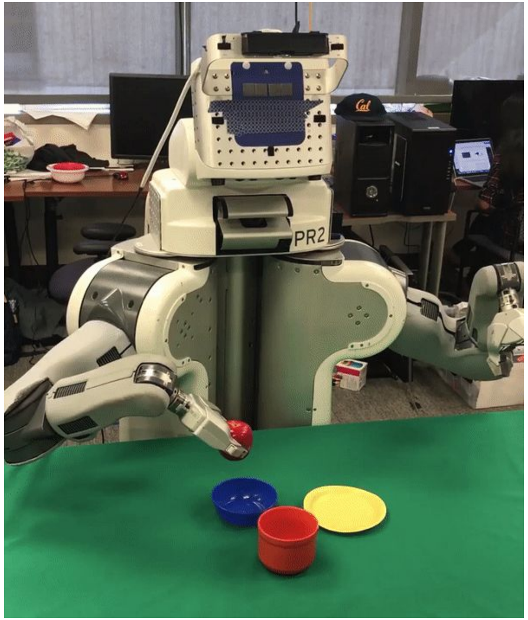
(Yu & Finn et al 2018)

Other works
Completing tasks
Diversity of objects

Our work
Closely mimicking motions
Diversity of motion
Completing tasks

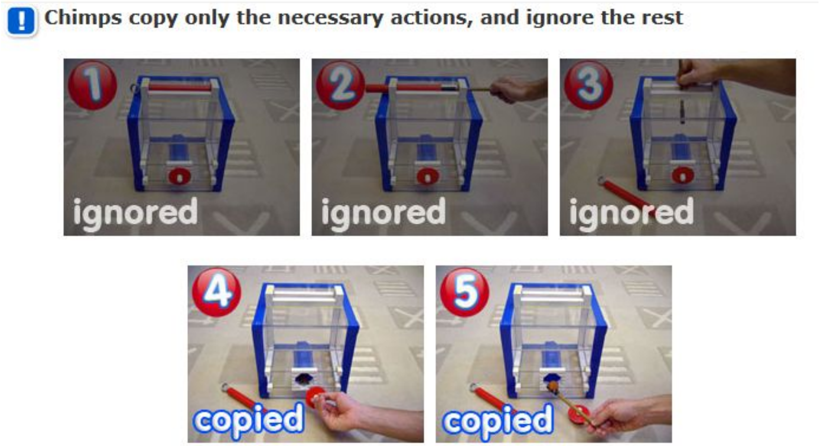


Demonstration



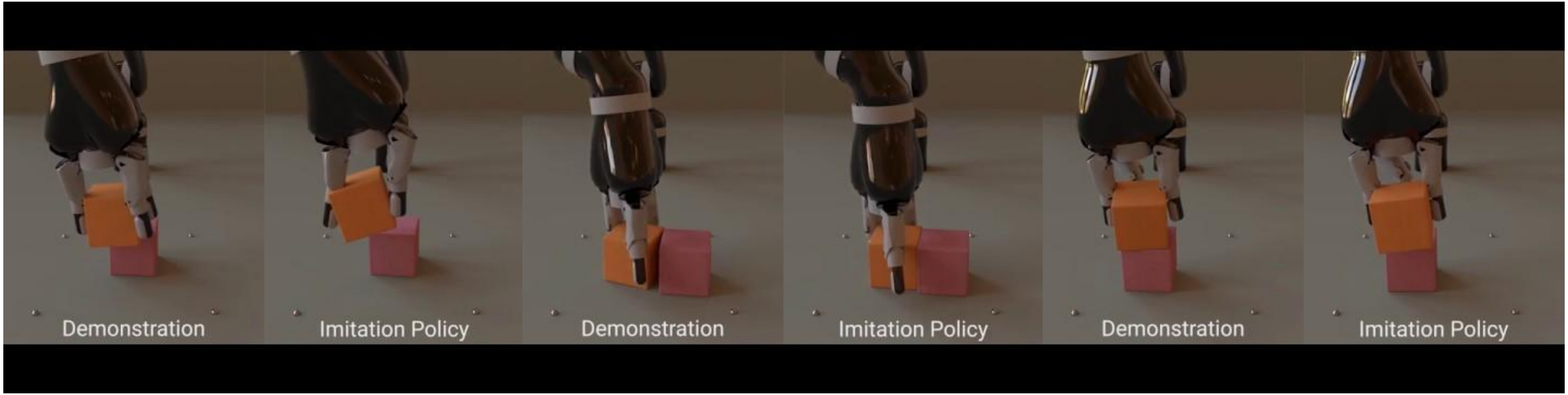
Policy

Over Imitation



MetaMimic: One-Shot High-Fidelity Imitation

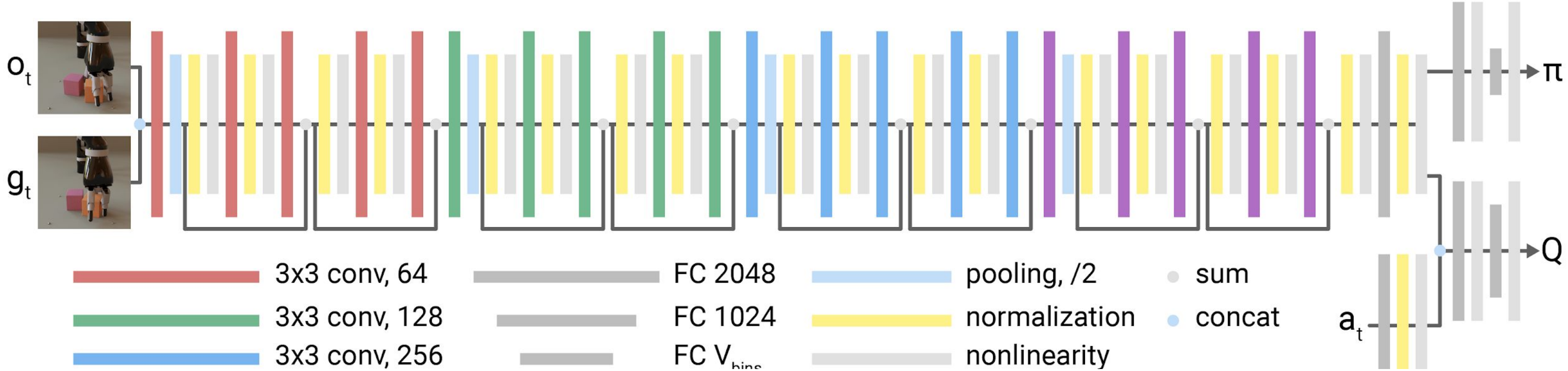
Imitation policy on
training demonstrations

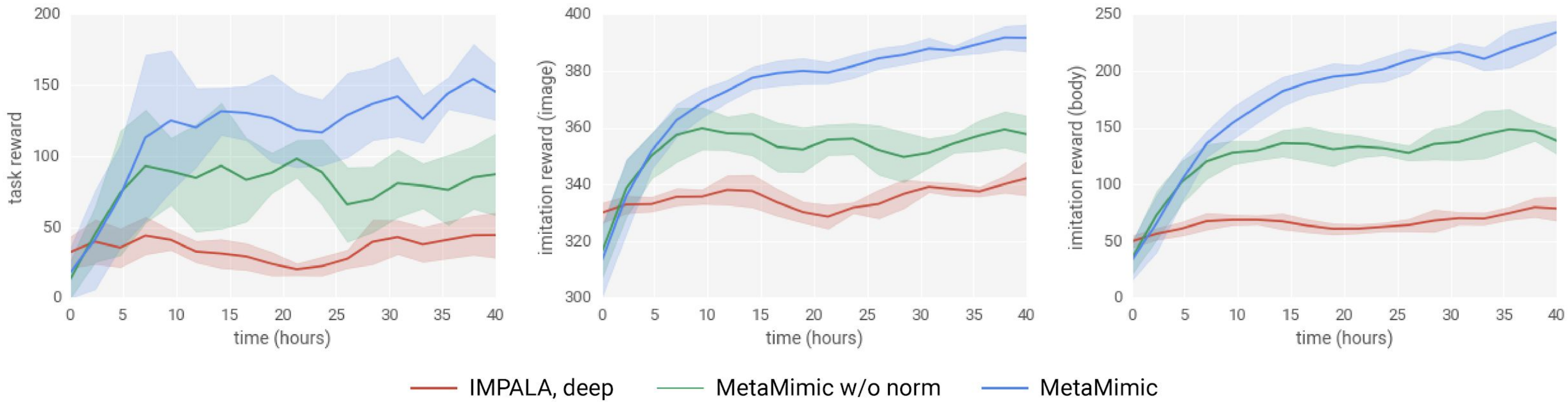


Important: Generalize to new trajectories

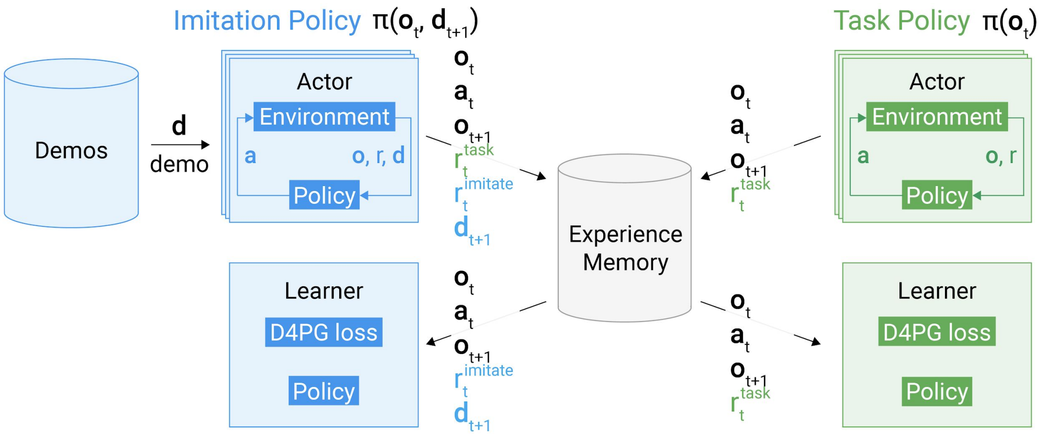
Imitation policy on
unseen demonstrations

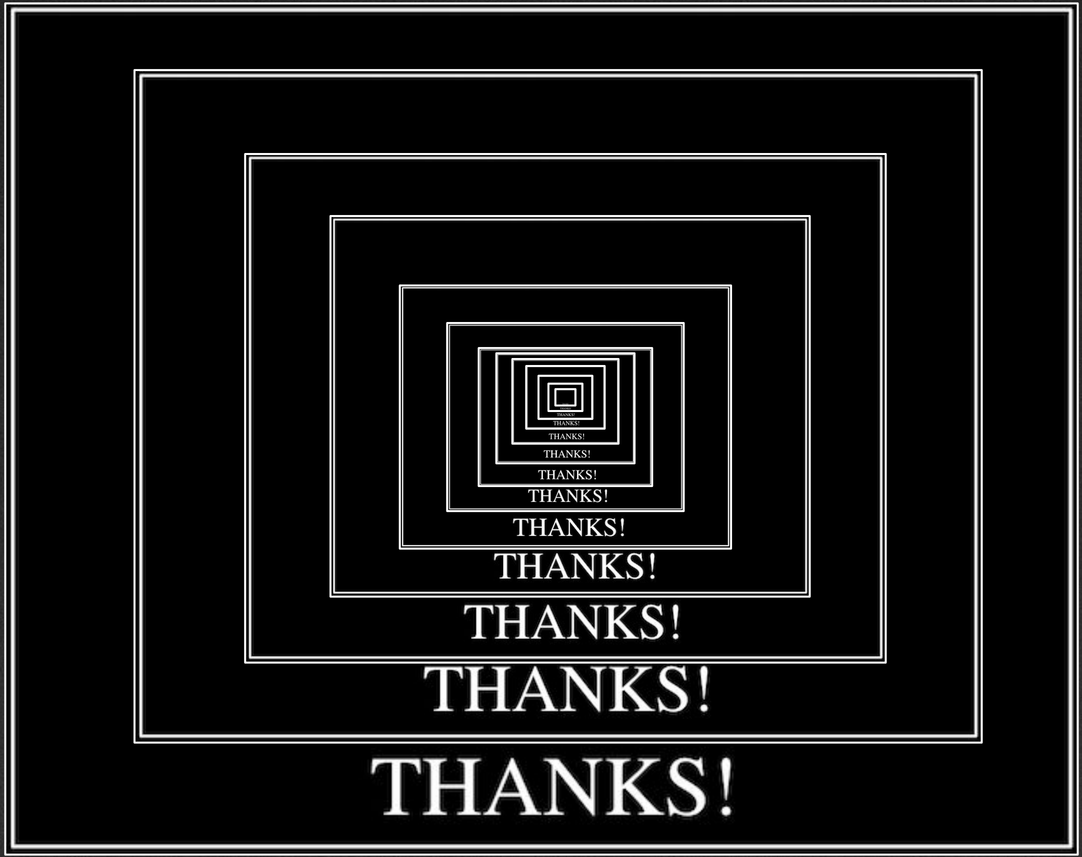
Massive deep nets are essential for generalization
And Yes!!! They can be trained with RL





MetaMimic Can Learn to Solve Tasks More Quickly Thanks to a Rich Replay Memory Obtained by High-Fidelity Imitation




THANKS!
THANKS!
THANKS!
THANKS!
THANKS!
THANKS!
THANKS!
THANKS!
THANKS!