Structure learning and the growth of skills

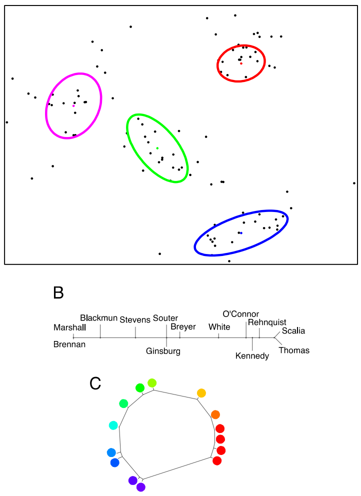
Anne Collins
Department of Psychology and Helen Wills Neuroscience Institute
University of California, Berkeley

Central problem of structure learning

What's out there?

How should I interact with it?

What is structure learning?



How many clusters?
How many features?
Which structural form?
Which functional form?

What is structure learning?

How many rules?
What is relevant?
Which learning pattern?

Interacting with the world

Markov Decision Process (MDP)

S - Set of States

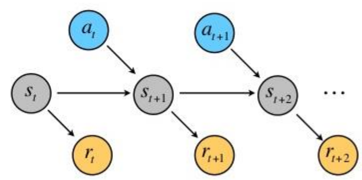
A - Set of Actions

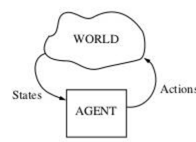
$\Pr(s' \mid a, s)$ - Transitions

α - Starting State Distribution

γ - Discount Factor

$r(s)$ - Reward [or $r(s,a)$]

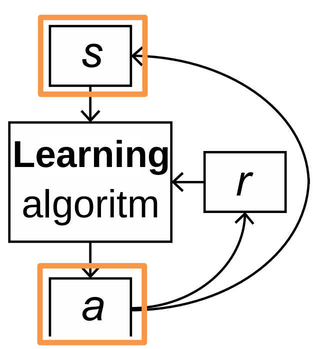




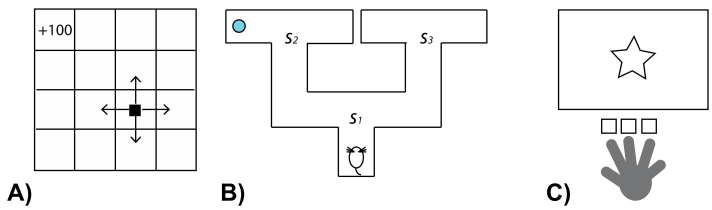
Policy π = P(a | s)

What are the inputs to the algorithm?

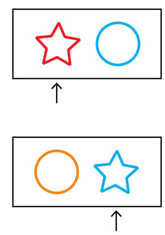
- How should I represent the state space?
- What is the relevant action space?
- What should my policy be?



What are states/actions?



A-C) Actions = motor movements



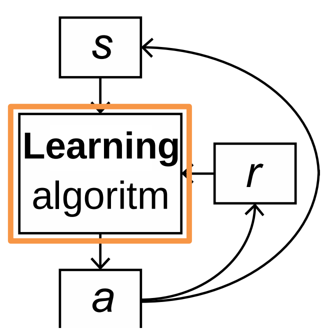
D) $\mathcal{A}$ = {“pick star”, “pick circle”}

~~$\mathcal{A}$ = {“pick blue”, “pick red”, …}~~

D)

Wilson & Niv 2011, …

With what algorithm should I learn?



Example: reinforcement learning

- Model-free RL:
  - $V_{t+1}(s_t) \leftarrow V_t(s_t) + \alpha\ (r_t+\gamma\ V_t(s_{t+1})\ -\ V_t(s_t))$
- Model-based RL:
  - Forward planning with a model of transitions
- Other:
  - Working Memory
  - Sampling from episodic memory
  - Bayesian hypothesis testing
  - …

The big picture

- We're still working to discover our hypothesis space.
- This space is over our interactions with the world, not over the world itself.
- Many similar principles, with different constraints.

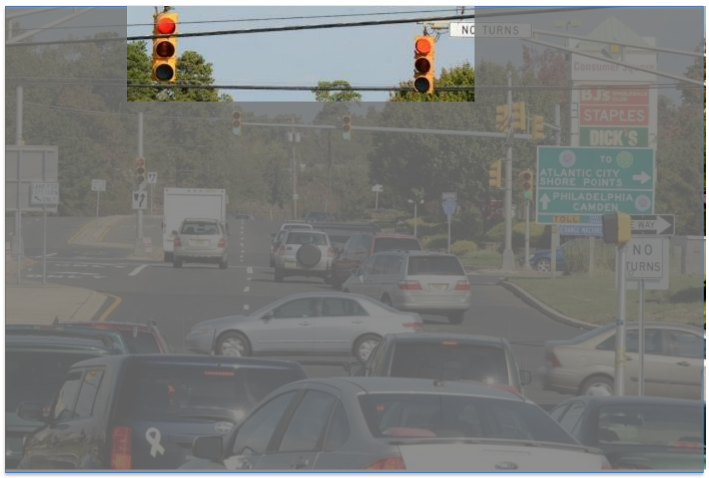
Building blocks

1. Structuring the inputs : state spaces
2. Structuring the outputs: action spaces
3. Structuring policies: hierarchy
4. Structuring learning: learning to learn

5. How does the brain do it?

PART 1: STATE SPACES

What is the state space?



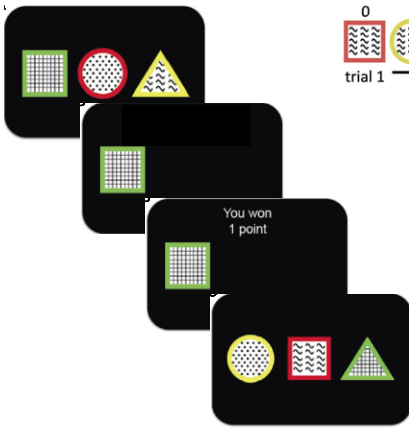
What is the state space?

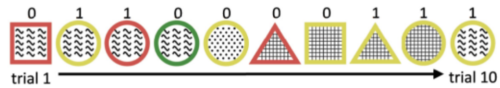
- Real life learning suffers from the curse of dimensionality

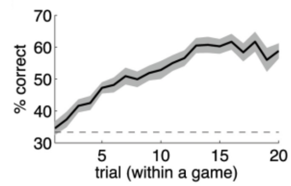
- Structure learning: Compressing the environment into a small, relevant state space

<u>Hypothesis</u>: Individuals structure the state space to represent only relevant information

Simplifying the representation of the state space







Wilson & Niv 2012
Niv et al 2015

Structure learning

Bayesian

- Hypothesis space:

  Which of nine features is predictive of reward?

Approximations

- Naïve RL

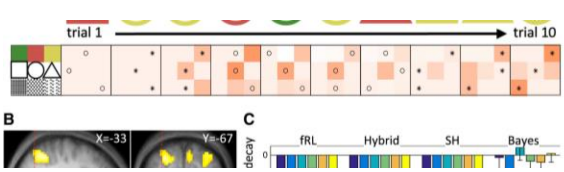
  learning for each 27 stimuli

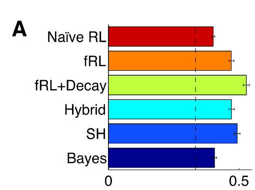
- Feature RL

  learning for each feature

- Hybrid:

  feature RL, with attentional weights from Bayesian inference





Wilson & Niv 2012
Niv et al 2015

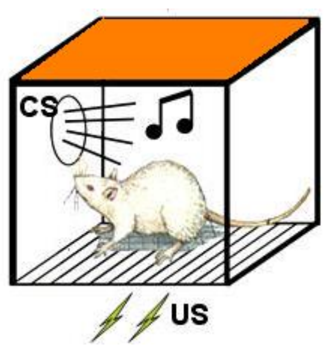
Structure learning: simplifying the problem

- Real life learning suffers from the curse of dimensionality
- By learning the structure of the state space, participants simplify the state representation and learn more efficiently
- This is better captured by approximate, attentional RL process than by optimal Bayesian inference

latent spaces - hypotheses

- States that are relevant for predicting outcomes may not be observable
- Structure learning may necessitate creating latent state spaces

What do animals learn during classical conditioning?



Slide from S Gershman

"It's that time of year when guys randomly explode."

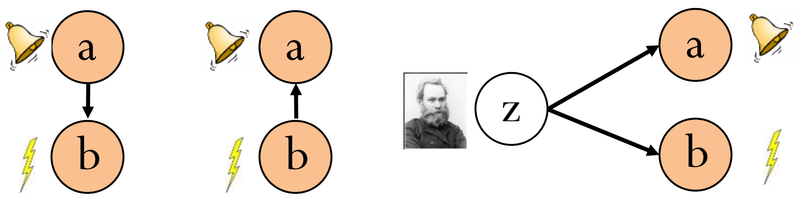
Slide from S Gershman

Some possibilities

Tone (a) causes shock (b)

Shock causes tone

Something else (z) causes both tone and shock

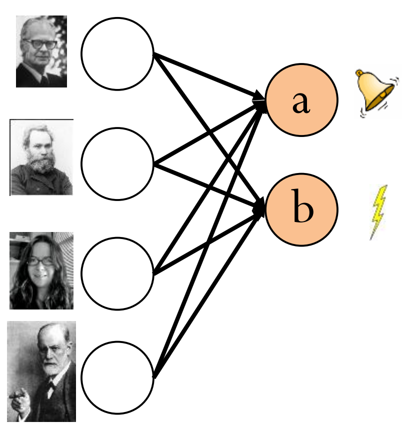


Too constrained

Too flexible?

Slide from S Gershman

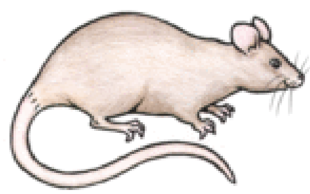
Too flexible?

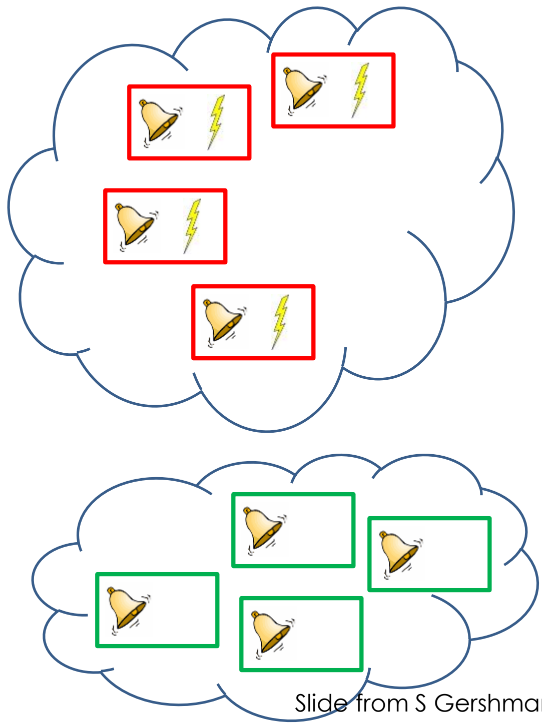


Hypothesis: Animals assume a generative model in which (1) the number of latent causes is unbounded, and (2) a small number of latent causes is more likely *a priori*.

Slide from S Gershman

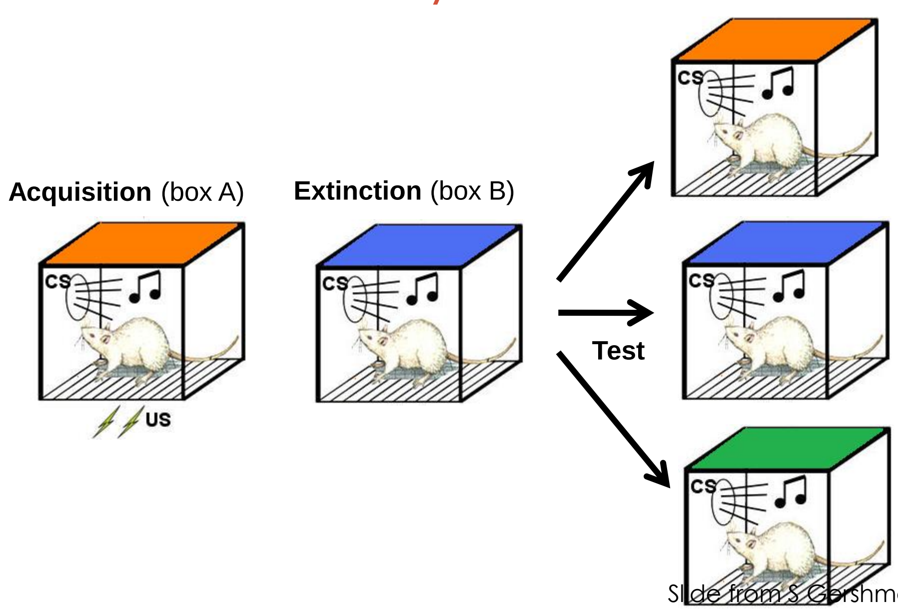
Conditioning as clustering





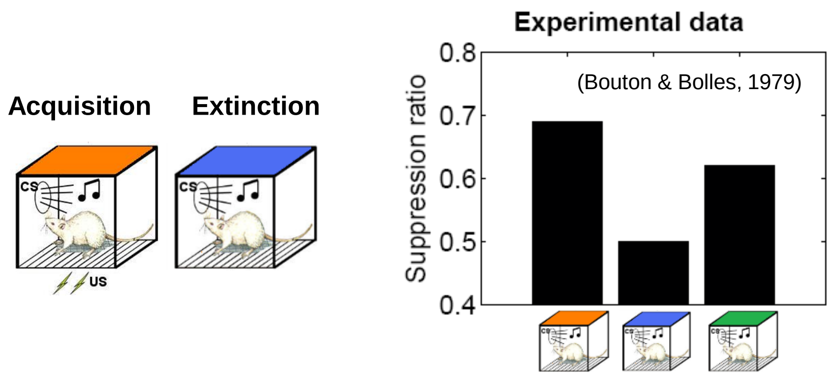
Slide from S Gershman

Case study: renewal



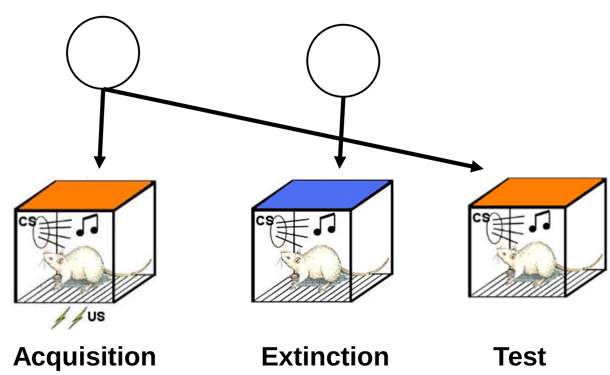
Slide from S Gershman

Conditioned responding is renewed!



The rat hasn't unlearned its conditioned response; it has ***learned something new.***

Slide from S Gershman

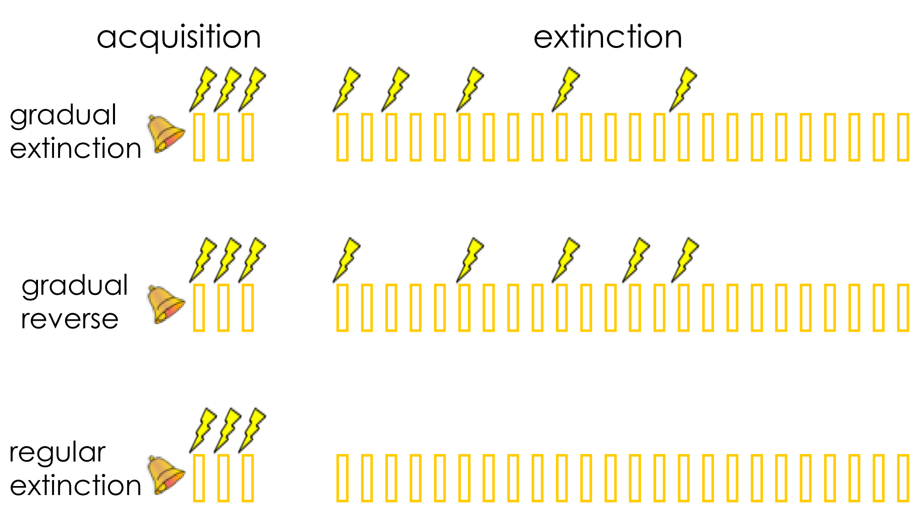


Slide from S Gershman

How to erase a fear memory

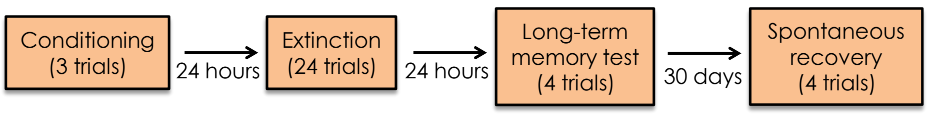
- If extinction induces inference of a new latent cause, we should be able to prevent the return of fear by tricking the brain into modifying the acquisition latent cause.
- We can do this by *extinguishing gradually*.

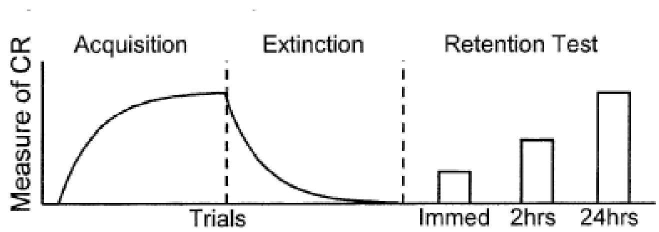
Slide from S Gershman



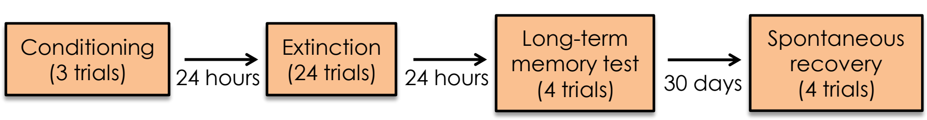
Slide from S Gershman

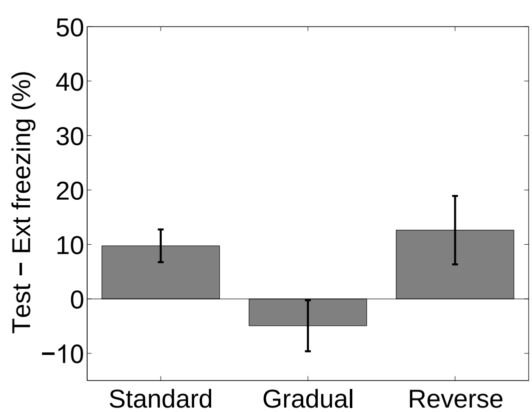
Experimental design





Slide from S Gershman





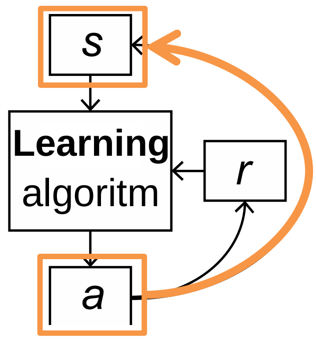
Gershman, Jones, Norman, Monfils & Niv (2013)

State spaces

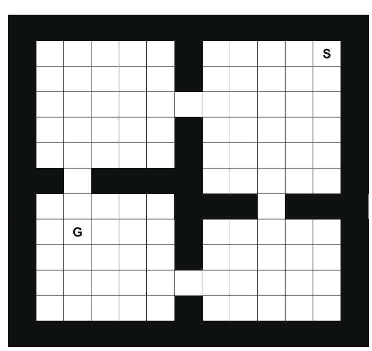
- Previous principles apply to learning the structure of the state space by clustering based on the predicted interactions with the environment
- Exact inference does not capture behavior well – approximate algorithms do better.

STRUCTURING ACTION SPACES

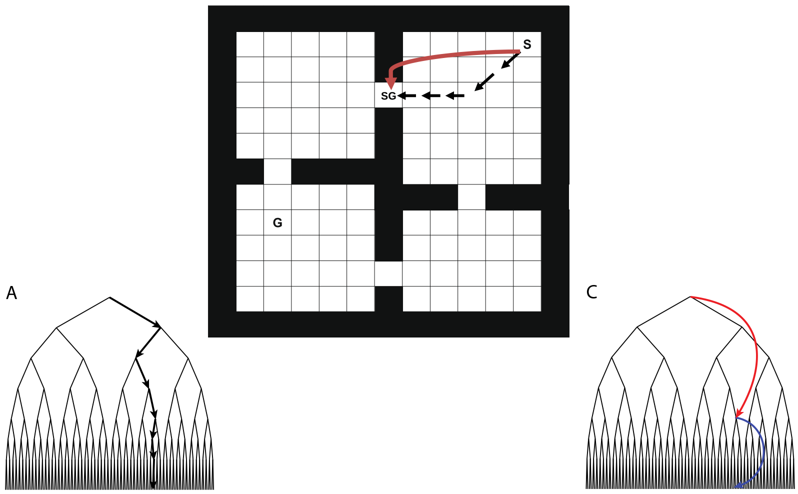
What are the inputs to the algorithm?



What is a good representation of the action space?

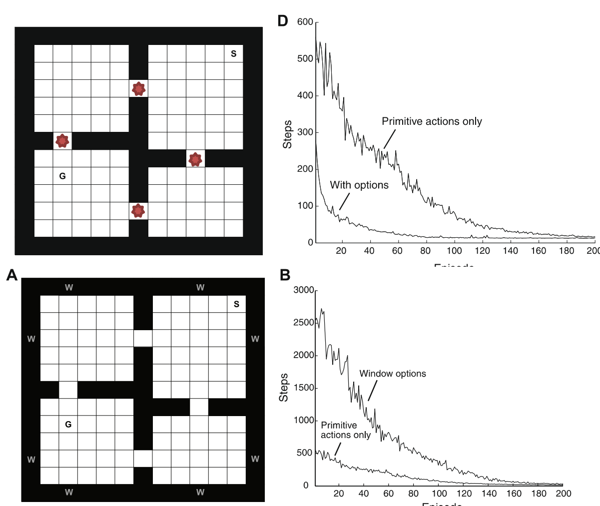


Exploration: options



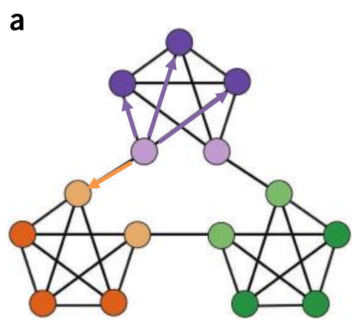
(Precup & Barto; Botvinick, Niv & Barto)

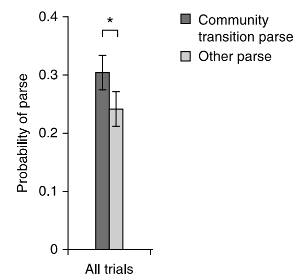
Learning the right option structure is critical



Botvinick, Niv & Barto; Solway et al, PCompBio 2014

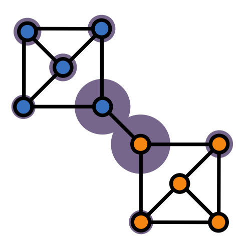
Learning the transition structure of the environment

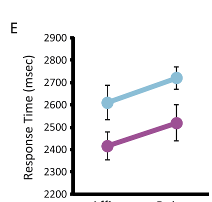




Schapiro et al 2013

Bottlenecks as optimal subgoals for hierarchical structure

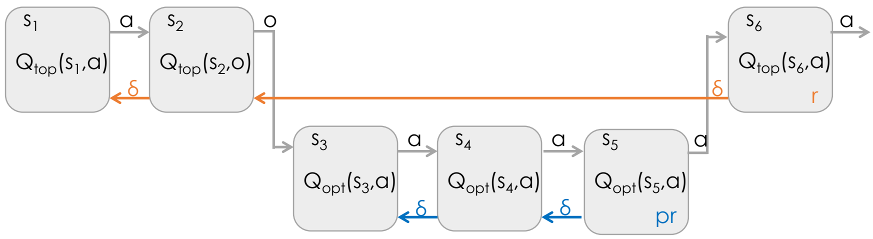




bottleneck
other

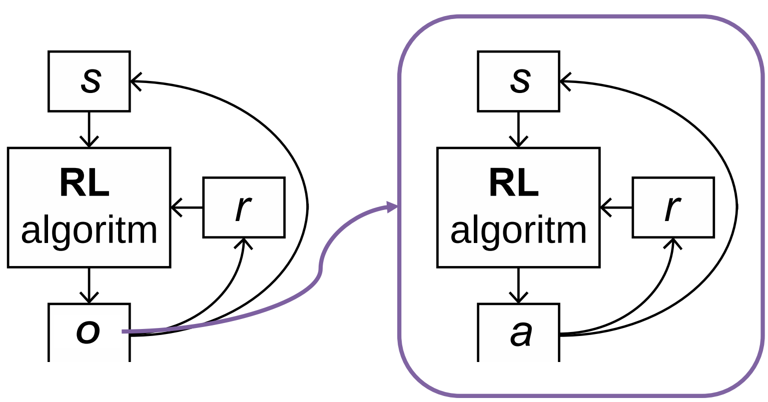
Solway et al; Diuk et al; Ribas-Fernandez et al.

Hierarchy in actions

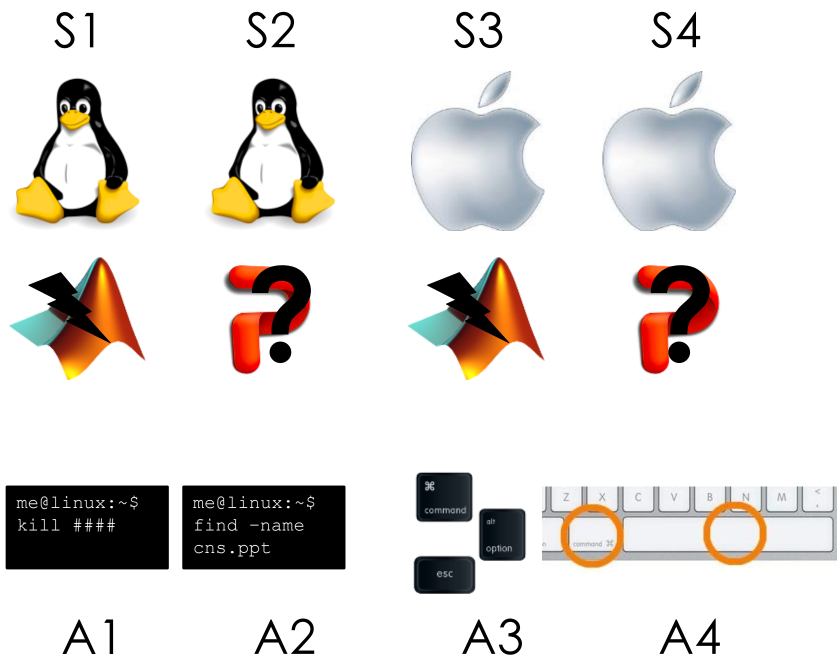


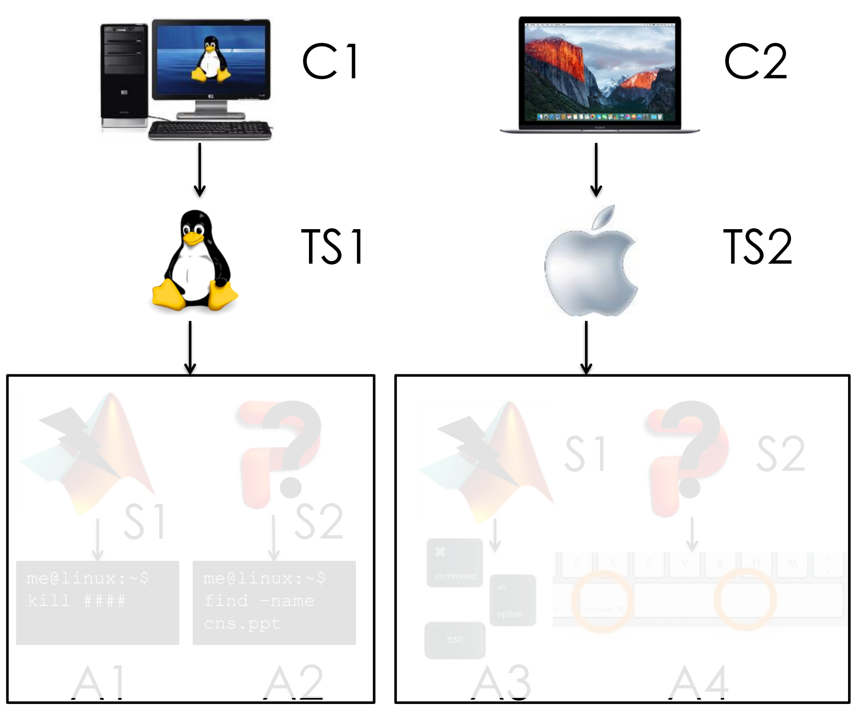
- Options: temporal hierarchy in action space
- Learning occurs in parallel at two hierarchical level

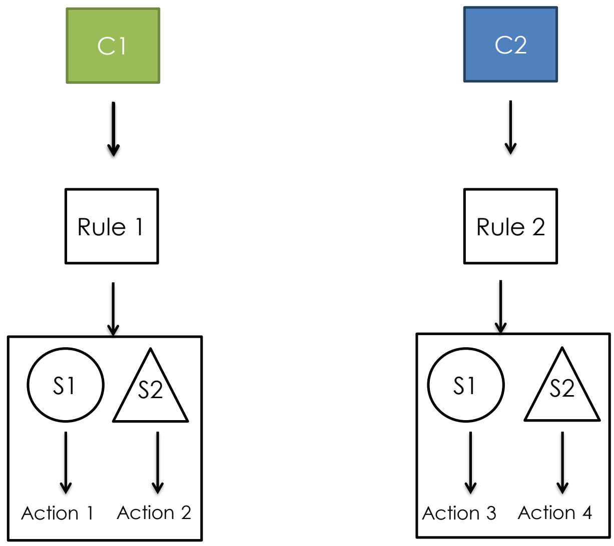
Hierarchical reinforcement learning



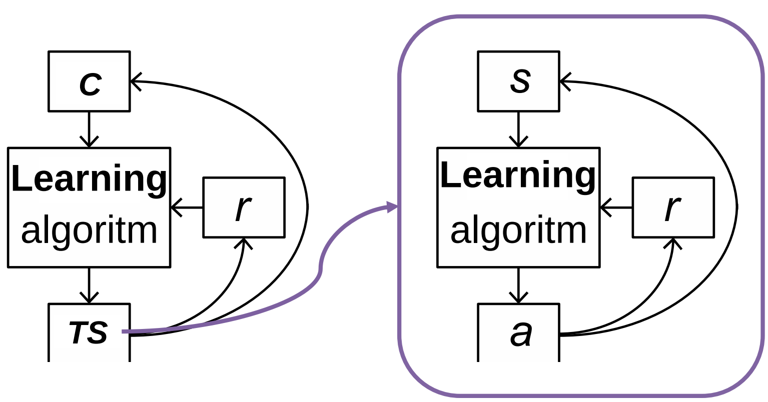
HIERARCHY


S1
S2
S3
S4
me@linux:~$
kill ####
me@linux:~$
find -name
cns.ppt
command
alt
option
esc
Z
X
C
V
B
N
M
A1
A2
A3
A4


C1
C2
TS1
TS2
S1
S2
me@linux:~$
kill ####
me@linux:~$
find -name
cns.ppt
A1
A2
S1
S2
A3
A4



Hierarchical reinforcement learning: levels of abstraction



parallel state and action states

- abstract states: contexts
- abstract actions: task-sets

Hierarchy in RL: learning over multiple state/action spaces

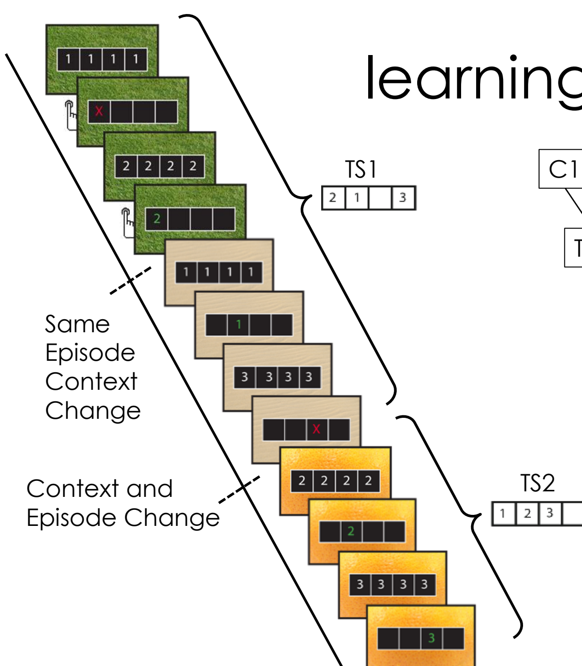
$$V(C,TS) \rightarrow \pi(c,TS) = p(TS \mid c)$$

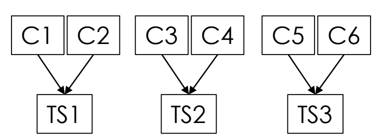
$$\downarrow$$

$$V_{TS}(s,a) \rightarrow \pi_{TS}(s,a) = p(a \mid s,TS)$$

$$V(s,a) \rightarrow \pi(s,a) = p(a \mid s)$$

learning task-sets

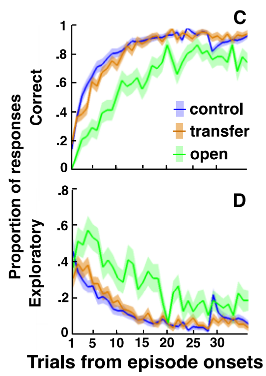


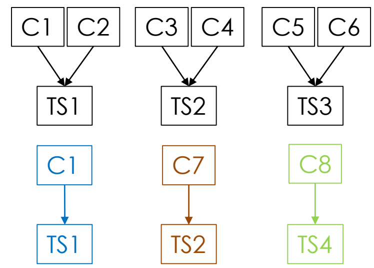


Learning
- non overlapping TS
- 25 episodes of 36-54 trials
- probabilistic feedback p=0.9

Collins & Koechlin, Plos Biology, 2012
Ekovich et al, in prep

Exploring with abstract actions: transfer of skills

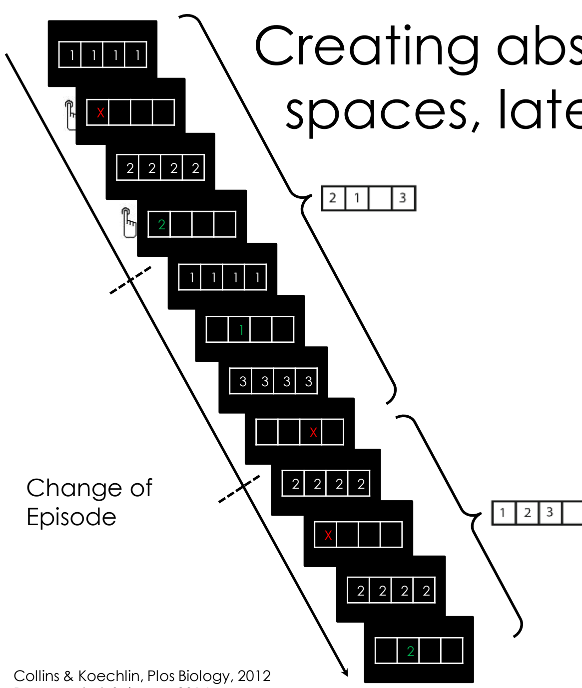


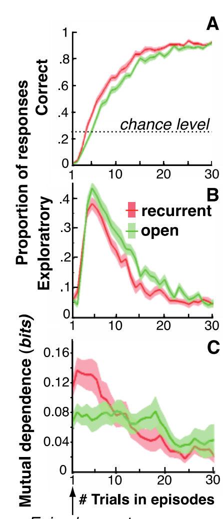


Testing **transfer** of abstract actions
- new context, old TS episodes
- new context, new TS episodes

Collins & Koechlin, Plos Biology, 2012
Ekovich et al, in prep

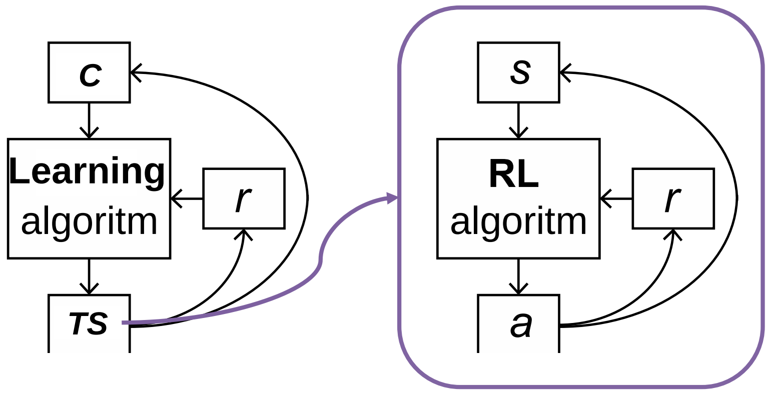
Creating abstract action spaces, latent context





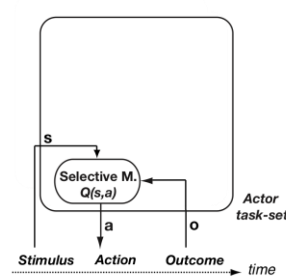
Collins & Koechlin, Plos Biology, 2012
Donoso et al, Science, 2014

learning task-sets with latent states



Reinforcement Learning (RL)

• Actor task-set continuously **adjusts** according to action **outcome** values



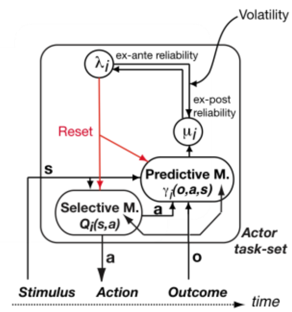
(Sutton & Barto, 1998; O'Doherty et al., 2004)

Uncertainty Monitoring (RL+UM): Change detection

• Reinforcement Learning and Monitoring **Uncertainty** of external contingencies and behavior **reliability**

• Actor task-set reliability, i.e. its ability to **predict** action outcomes, is **inferred online** (Bayesian inference)

• The **actor task-set** is reset whenever it becomes unreliable



(Yu & Dayan, 2005; Behrens et al., 2007)

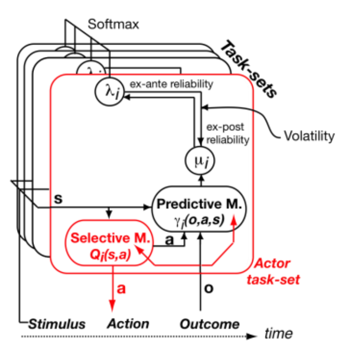
Multiple RL+UM

optimally tracking a fixed number of hypotheses

• Reinforcement Learning and Monitoring:
1. Uncertainty
2. Reliability of **multiple alternative task-sets**

• Relative reliabilities of a **fixed** collection of concurrent task-sets inferred online

• Actor task-set selected based on **reliability**



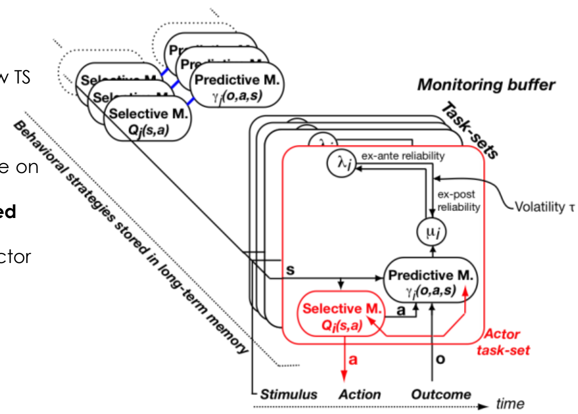
(Doya & Kawato, 2002; Samejima & Doya, 2007)

(RL+UM)+PROBE
approximately tracking an unknown number of hypotheses

• Reinforcement learning and Monitoring:
1. **Uncertainty**
2. **Reliability** of multiple alternative TS
3. Opportunity to **create** new TS

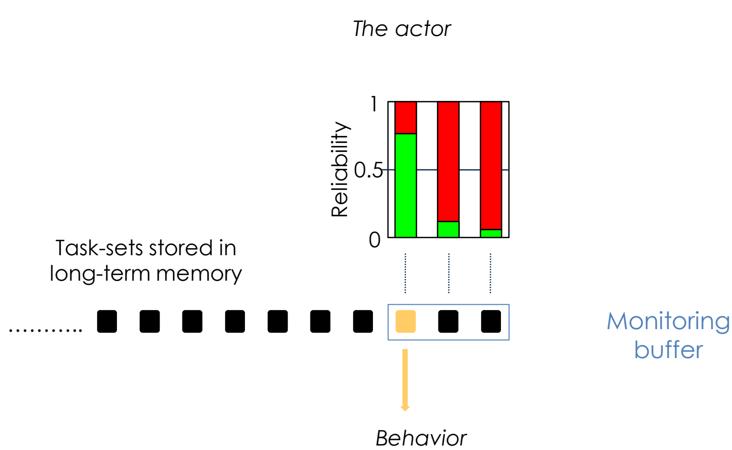
• **TS creation** obeys 2 constraints:
- Forward, online Bayesian inference on TS reliability
- Number of monitored TS is **bounded**

• How does the model select the actor task-set?

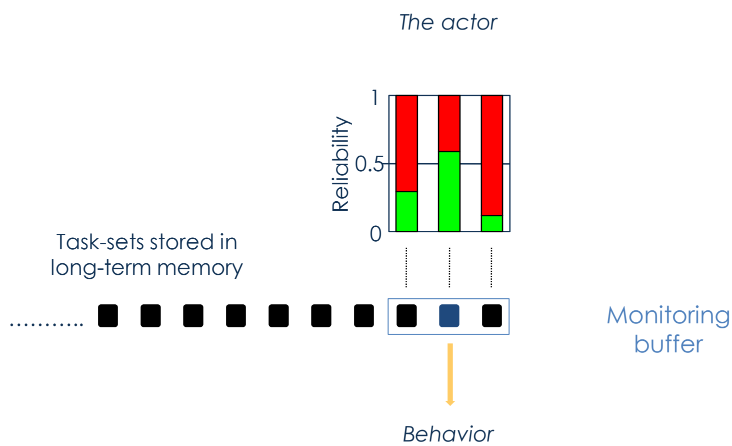


(Collins & Koechlin, PloS Biol, 2012)

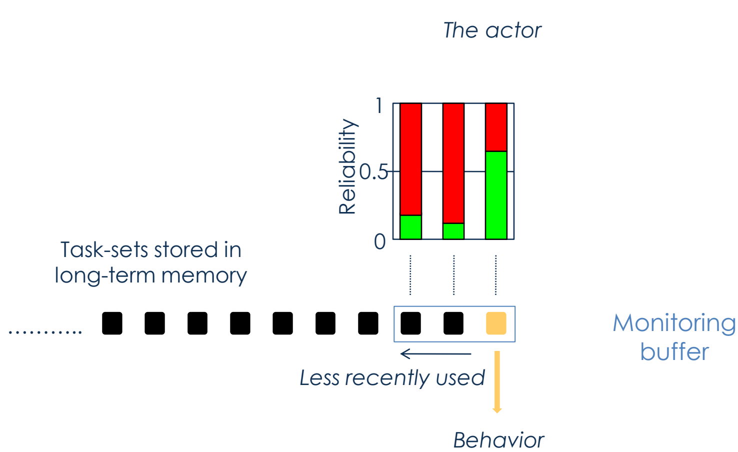
Exploitation

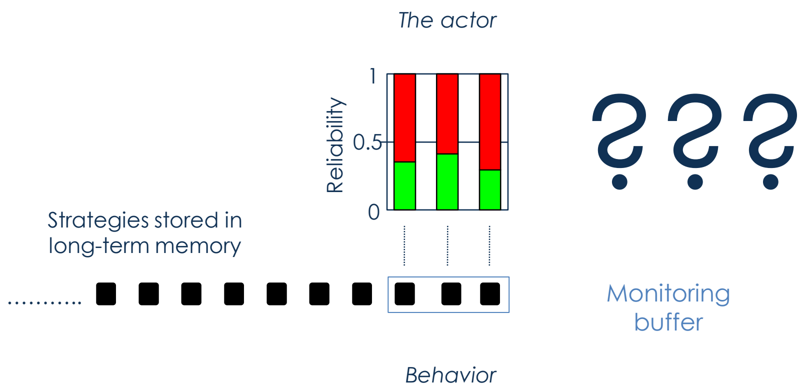


Exploitation

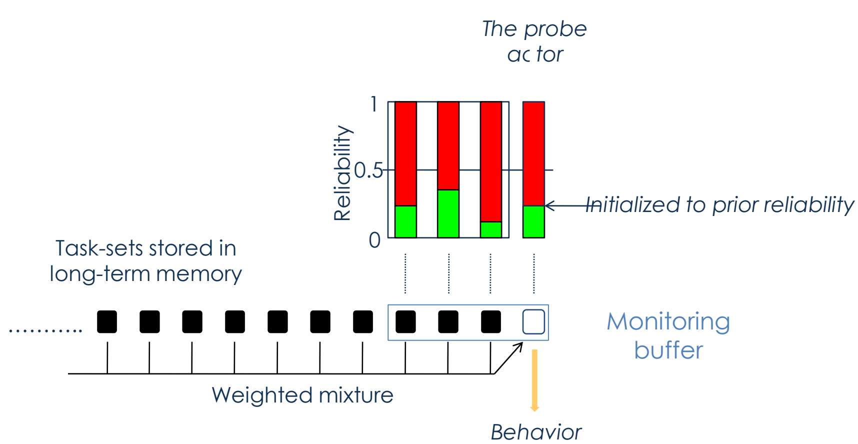


Exploitation

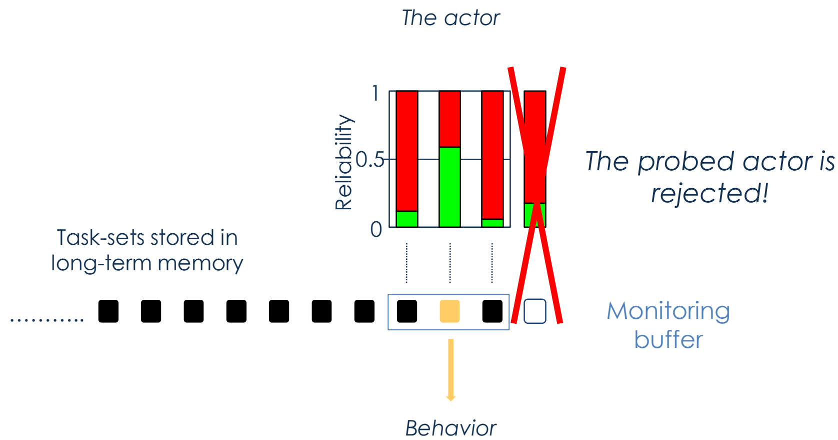




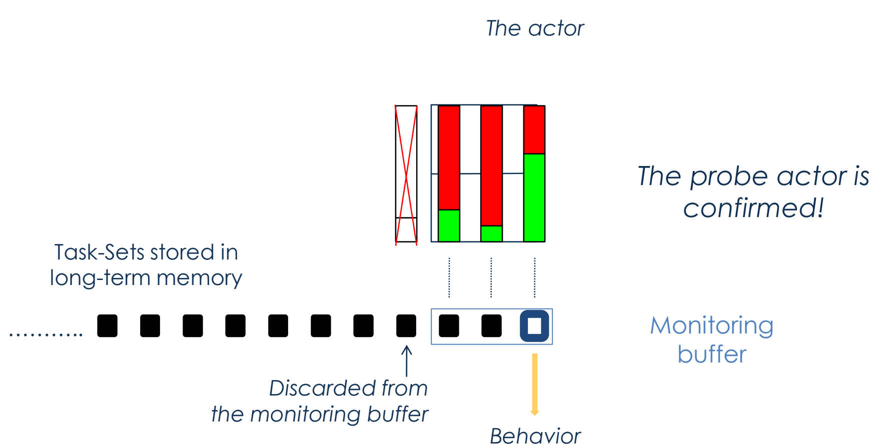
Switch from Exploitation to Exploration



Return to Exploitation (*Rejection* events)

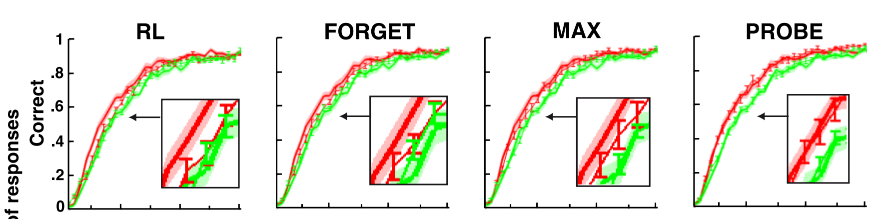


Return to Exploitation (*Confirmation* events)

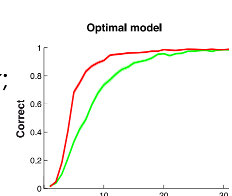


CRP-like clustering

- Contexts cluster together based **on environment contingencies**: stimulus-action-outcome mapping similarity
- Clusters index **TS rules**
  - provide ability to generalize TS to new context
  - ability to create new TS as needed
- Inference with **approximate** tracking of uncertainty over an unbounded hypothesis space: abstract task-sets
- Proposed algorithm defines discrete **high-level exploitation/exploration** periods.



- Probe model captures behavior best:
- Ability to
  - "probe" the need to create a new cluster;
  - monitor a small number of other hypotheses;
  - minimize default computational cost.

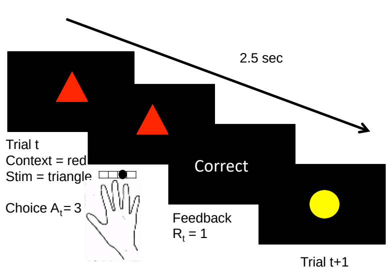


Learning task-sets

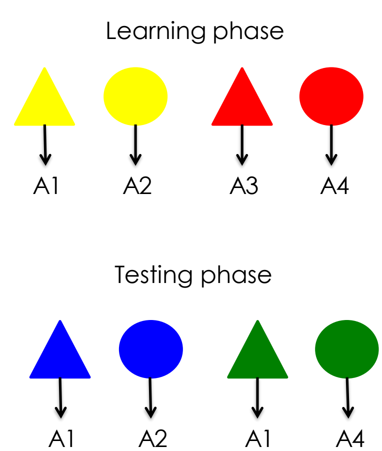
- Temporal stability makes TS structure useful:
  - by default, exploitation of current TS
  - only tracks complexity when decrease in reliability signals a need for control

- Does structure learning happen without such pressure?

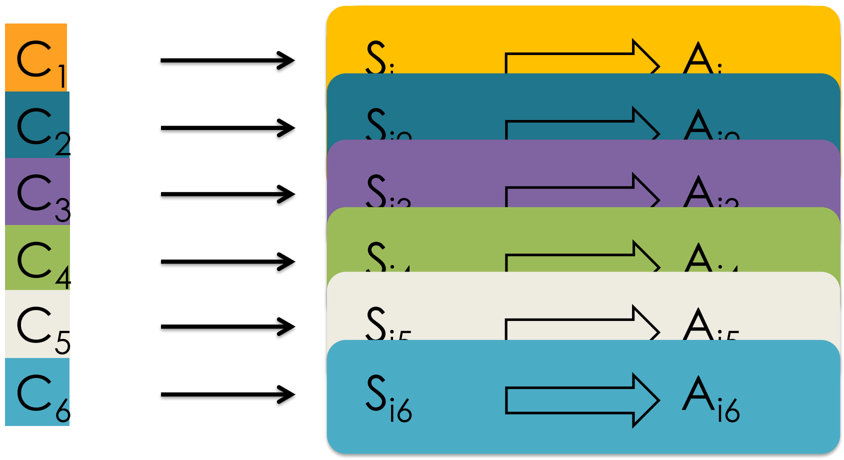
Hierarchical Structure learning occurs by default



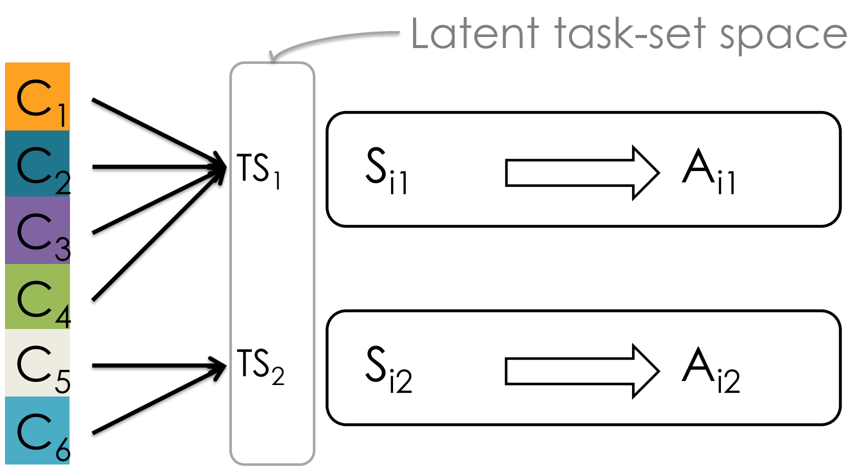
Collins & Frank, Psych Review, 2013



Computational Model

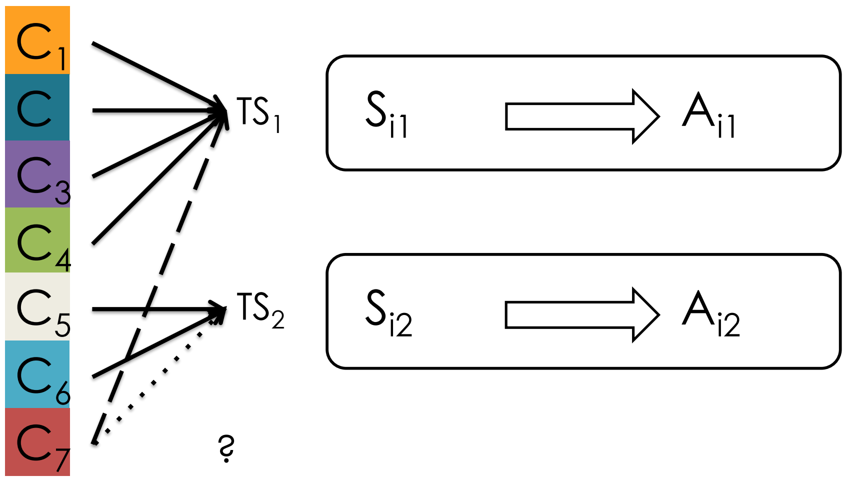


Abstracting Task-set rules

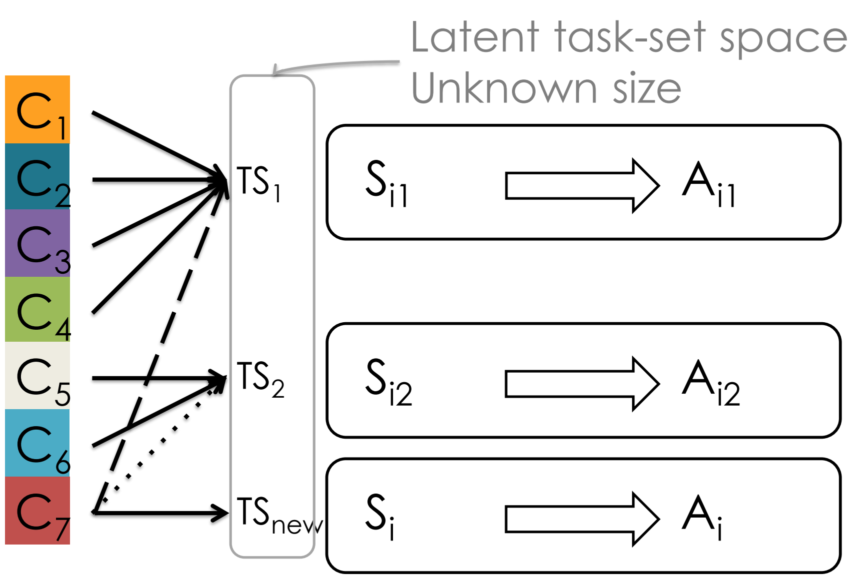


TS as abstract rule objects
Reverberi et al 2011
Woolgar et al 2011

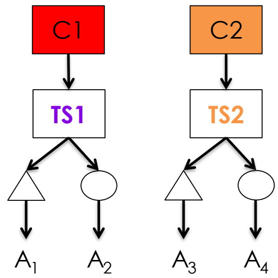
CRP prior on Task-set rules

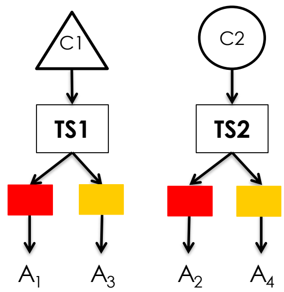


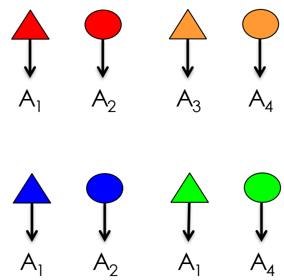
Ability to create new Task-sets

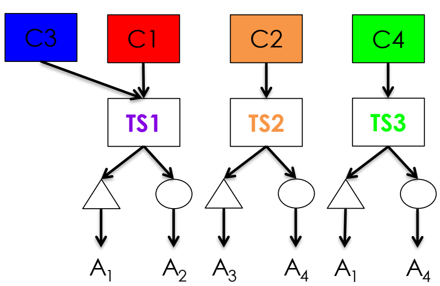


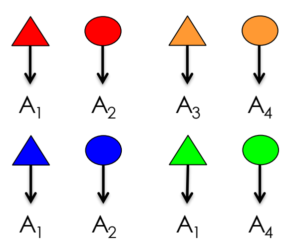
Collins & Frank, Psych Review, 2013

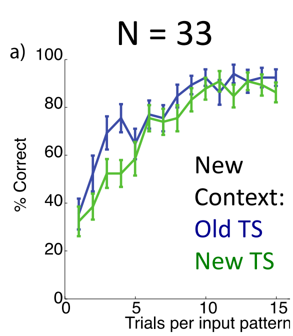


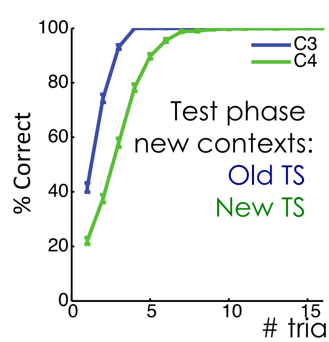








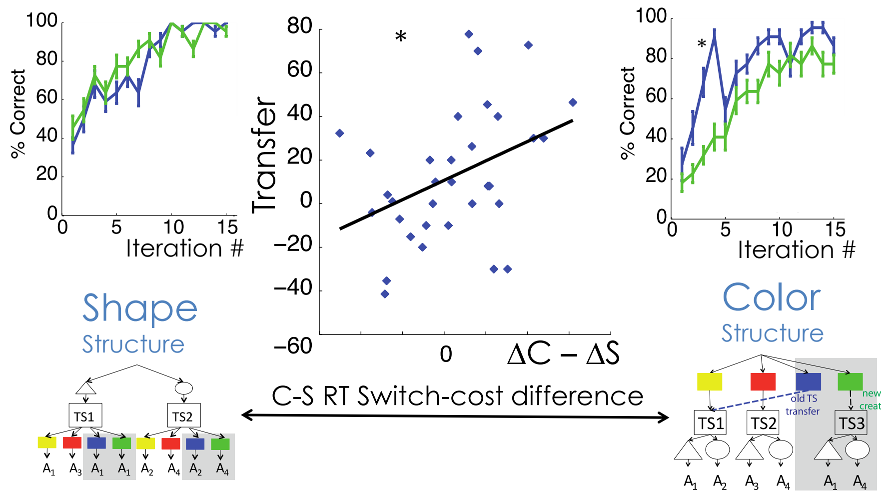




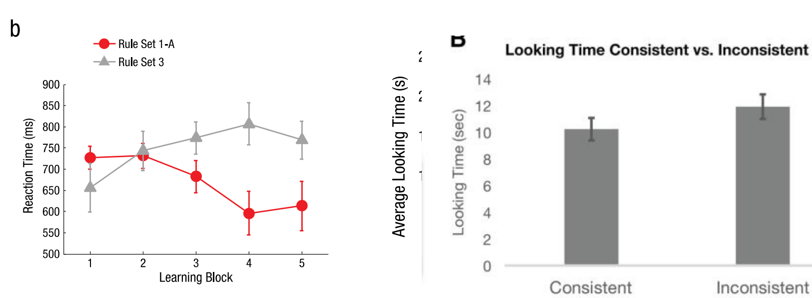
Positive transfer: In new context, faster learning for an old TS rule

Collins & Frank, Psych Review, 2013

Training phase RT switch-cost predicts phase 2 transfer



Structure learning and generalization in infants

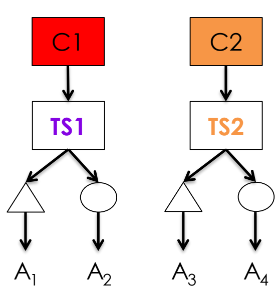


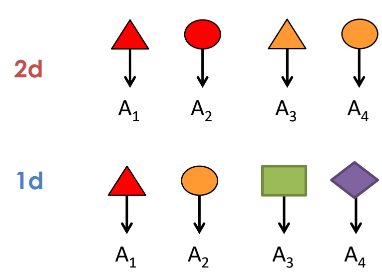
Werchan et al 2015, Psych Science
Werchan et al 2016, JoN

- Humans create
  - multiple state spaces (stimuli, contexts)
  - multiple action spaces (actions, task-sets)
  - at multiple hierarchical abstraction levels

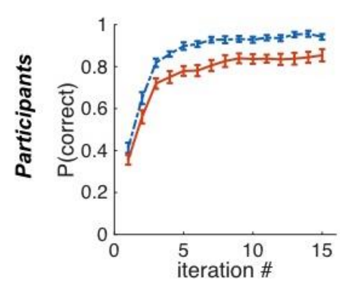
- This is a default behavior, even with no immediate gain

Learning structure is costly.

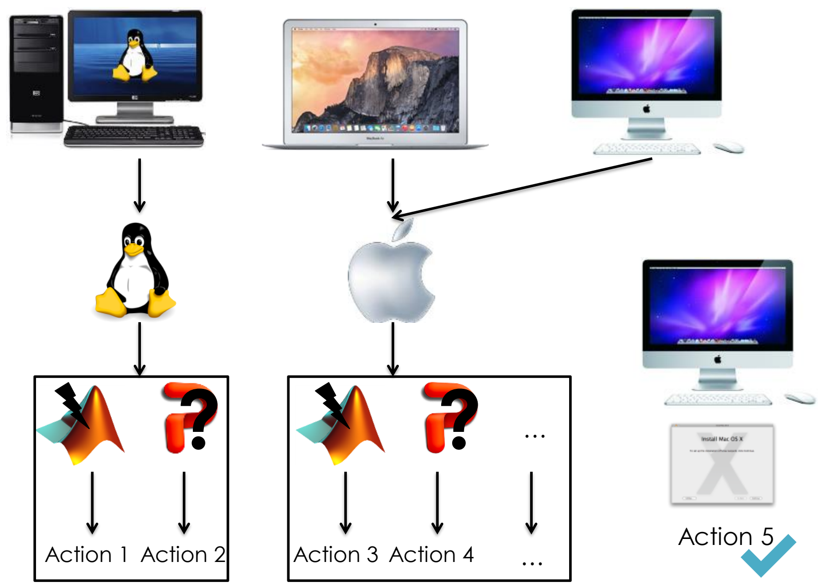


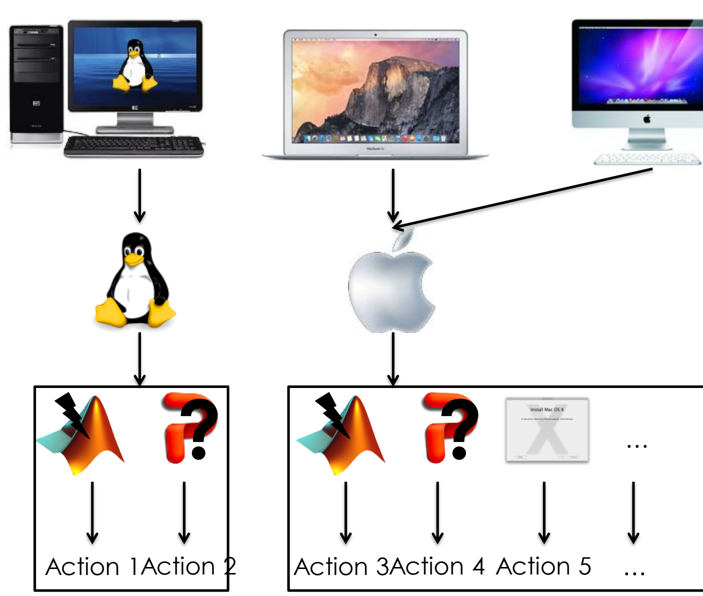


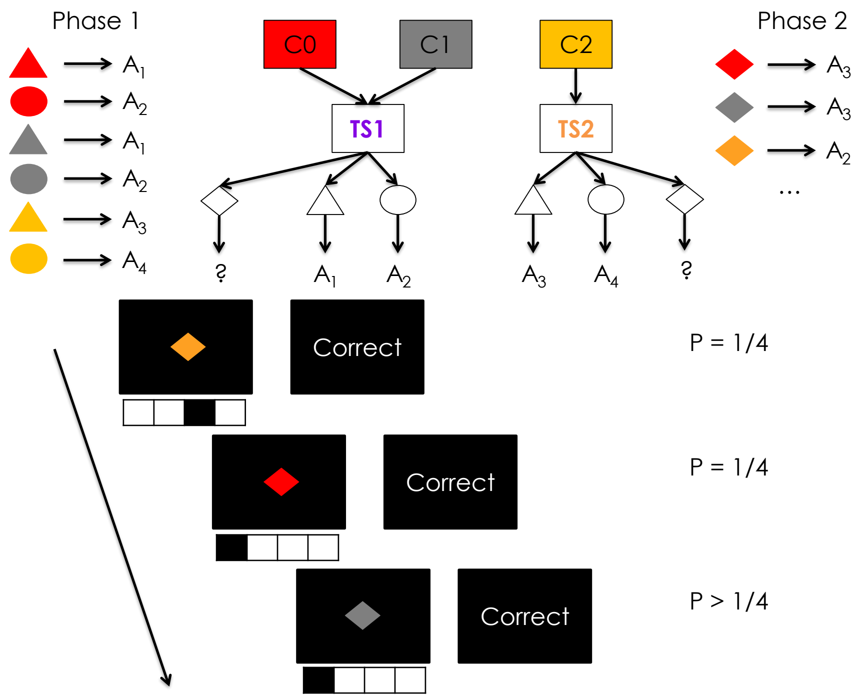
Cost is overriden by strong prior that structure learning is **long-term** beneficial



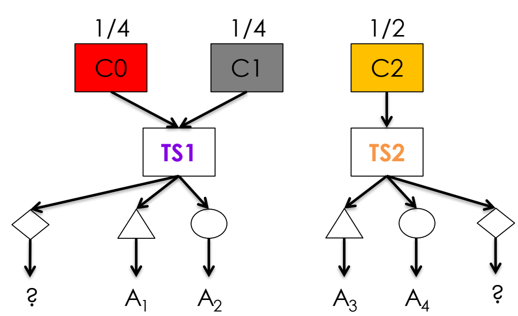
Collins, JoCN, 2017


Action 1
Action 2
Action 3
Action 4
...
...
Action 5

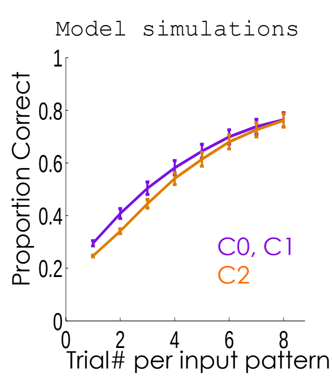

Action 1
Action 2
Action 3
Action 4
Action 5
...
...



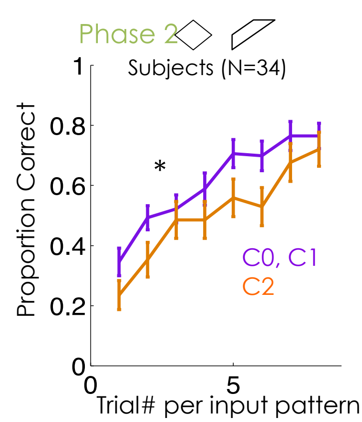
Phase 1
A1
A2
A1
A2
A3
A4
C0
C1
C2
TS1
TS2
?
A1
A2
A3
A4
?
Phase 2
A3
A3
A2
...
Correct
P = 1/4
Correct
P = 1/4
Correct
P > 1/4



1/4
1/4
1/2
C0
C1
C2
TS1
TS2
?
A1
A2
A3
A4
?

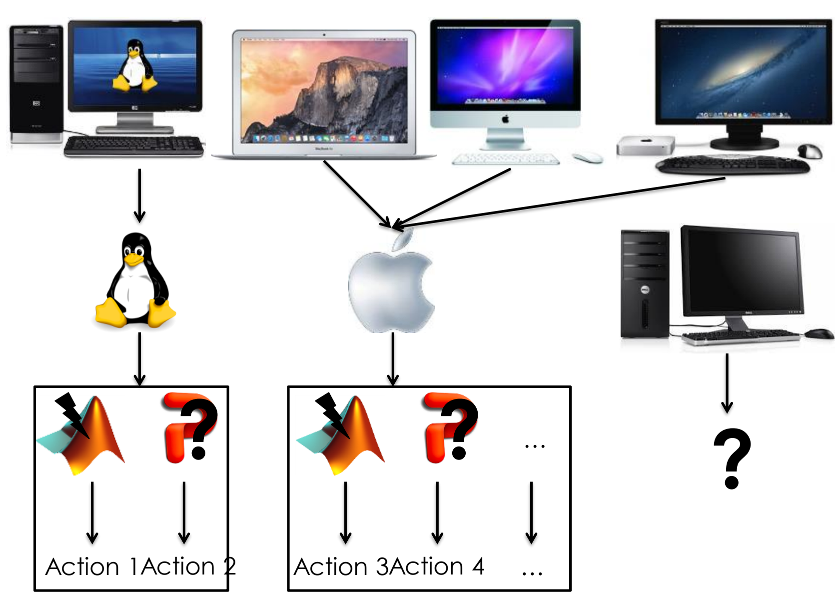


Model simulations
Proportion Correct
Trial# per input pattern
C0, C1
C2

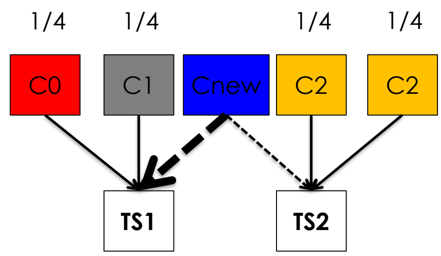


Phase 2
Subjects (N=34)
Proportion Correct
Trial# per input pattern
C0, C1
C2

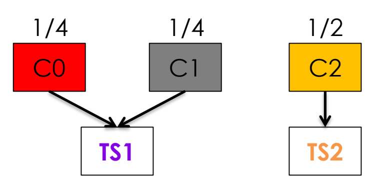
Structure learning enables immediate transfer of new learned skills

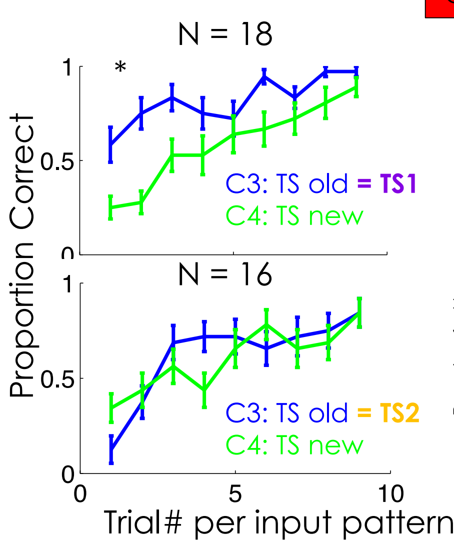


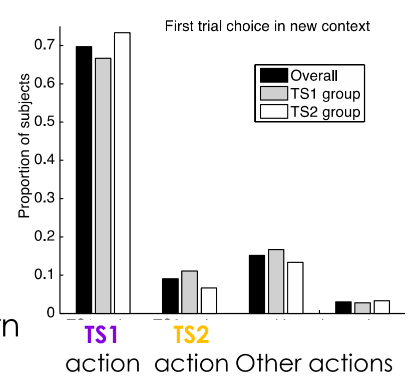
What rules do we explore in a new context?



CTS model predicts better transfer for C3 →TS1 than for C3→TS2







Collins & Frank, Cognition, accepted

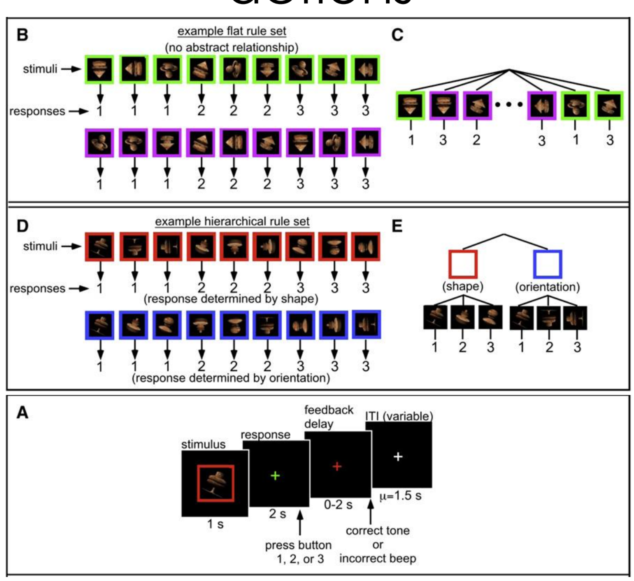
Subjects' generalization prior is stronger for more popular rules

* Consistent with model's context-popularity prior
* Not with a trial-frequency popularity prior

Structure learning: task-sets

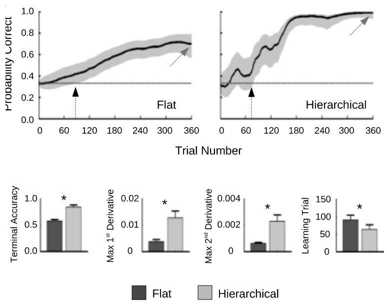
- TS learning is an example of **hierarchical** structure learning, with:
  - multiple states and action stats
  - abstract, latent context space
  - clustering that promotes generalization by fast, high-level exploration
- It exemplifies the fact that structure learning is a **default** behavior despite being costly
- It is best accounted for by approximations of rational non-parametric inference schemes

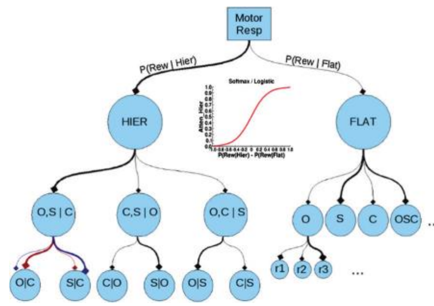
Hierarchical states, internal actions

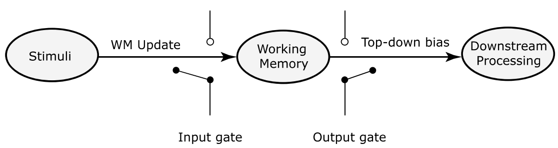


Badre et al, Neuron, 2010

Hierarchical states, internal actions

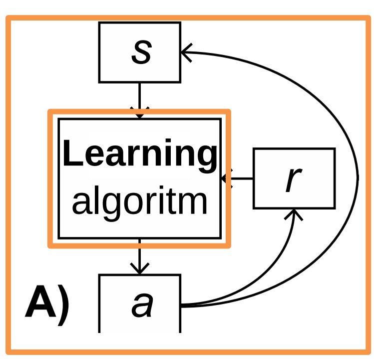






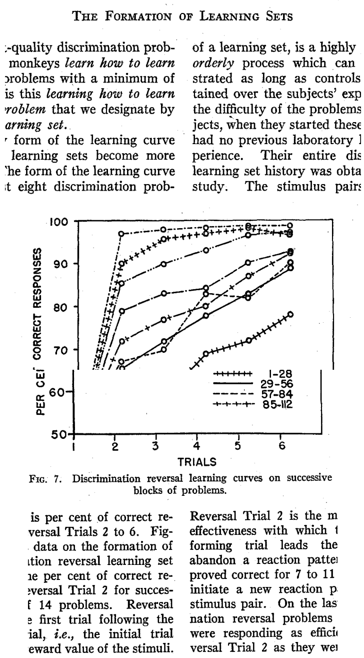
LEARNING TO LEARN

Learning to learn



Harlow 1949

- Monkeys and children


THE FORMATION OF LEARNING SETS

:-quality discrimination prob- monkeys *learn how to learn* problems with a minimum of is this *learning how to learn* roblem that we designate by arning set.

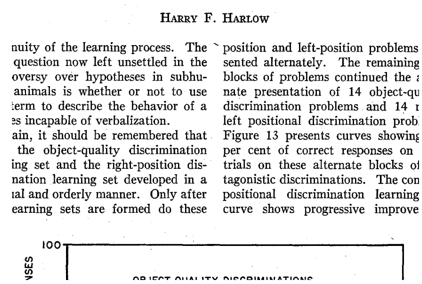
form of the learning curve learning sets become more he form of the learning curve t eight discrimination prob-

of a learning set, is a highly *orderly* process which can strated as long as controls tained over the subjects' exp the difficulty of the problems jects, when they started these had no previous laboratory perience. Their entire dis learning set history was obta study. The stimulus pairs

FIG. 7. Discrimination reversal learning curves on successive blocks of problems.

is per cent of correct re- versal Trials 2 to 6. Fig- data on the formation of ition reversal learning set e per cent of correct re- eversal Trial 2 for succes- 14 problems. Reversal e first trial following the ial, *i.e.*, the initial trial eward value of the stimuli.

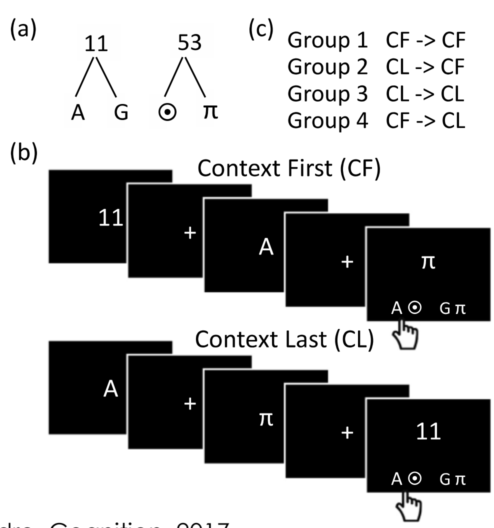
Reversal Trial 2 is the m effectiveness with which forming trial leads the abandon a reaction patte proved correct for 7 to 11 initiate a new reaction p stimulus pair. On the las nation reversal problems were responding as effici versal Trial 2 as they we


HARRY F. HARLOW

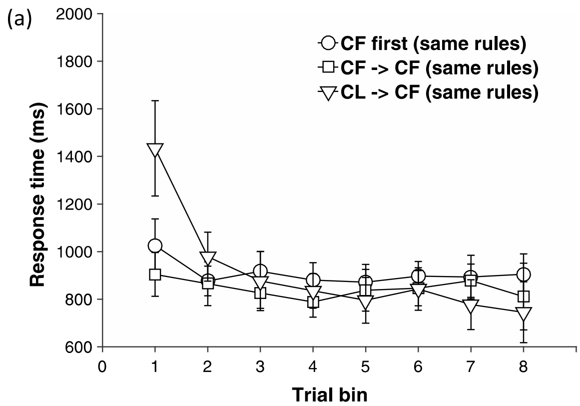
nuity of the learning process. The question now left unsettled in the oversy over hypotheses in subhu- animals is whether or not to use term to describe the behavior of a es incapable of verbalization. ain, it should be remembered that the object-quality discrimination ing set and the right-position dis- nation learning set developed in a ial and orderly manner. Only after earning sets are formed do these

position and left-position problems sented alternately. The remaining blocks of problems continued the a nate presentation of 14 object-qu discrimination problems and 14 left positional discrimination prob Figure 13 presents curves showing per cent of correct responses on trials on these alternate blocks of tagonistic discriminations. The con positional discrimination learning curve shows progressive improve

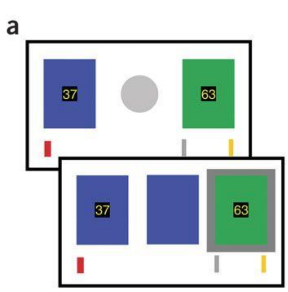
Learning the structure of a task



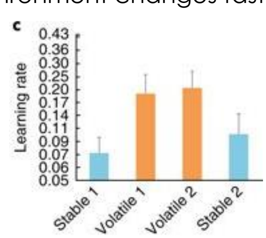
Bhandari & Badre, Cognition, 2017

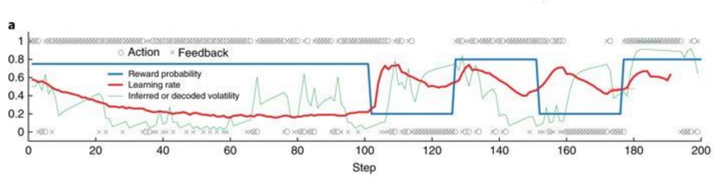


Learning the parameters of the learning algorithm



Learning rate increases when the environment changes faster





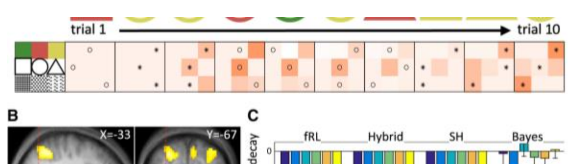
Building blocks

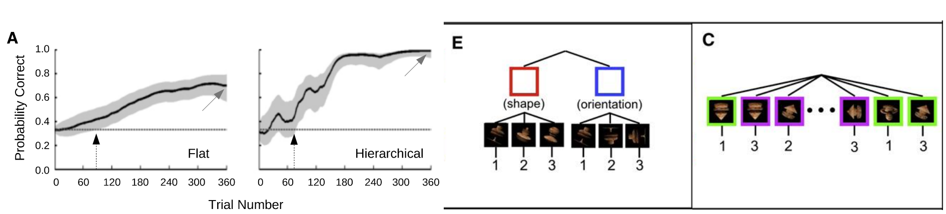
1. Structuring the inputs : state spaces
2. Structuring the outputs: action spaces
3. Structuring policies: hierarchy
4. Structuring learning: learning to learn

Structure learning

- Simplify → reduce curse of dimensionality
- Generalize → learn more flexibly
- Explore → learn more faster
- Adjust → learn more efficiently

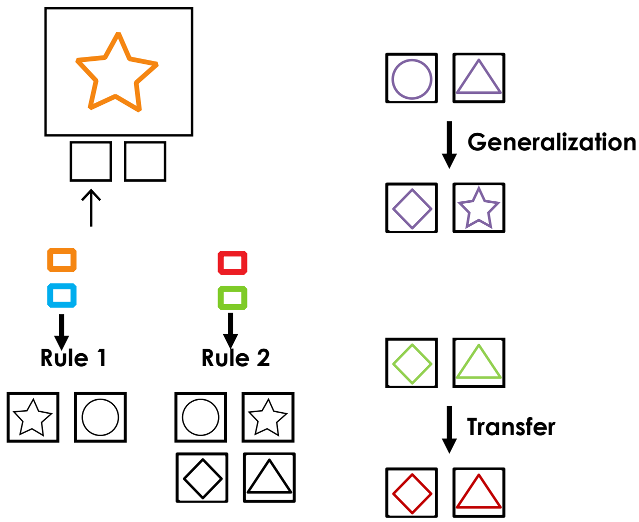
Simplifying the problem

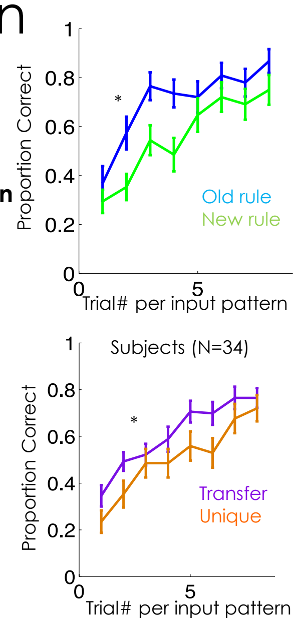




Badre et al 2010, Frank & Badre 2011

Abstract rules: transfer and generalization



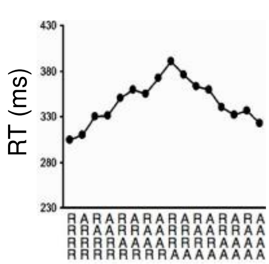


Collins & Frank 2012, 2016

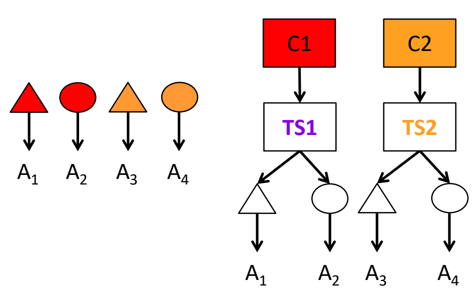
The structure learning bias

Complexifying the problem can lead to later simplification:

- creating more abstract, complex representations of the problem leads to more flexibility in their use
  - latent states and more abstract actions
- structure bias:



Yu & Cohen 2009



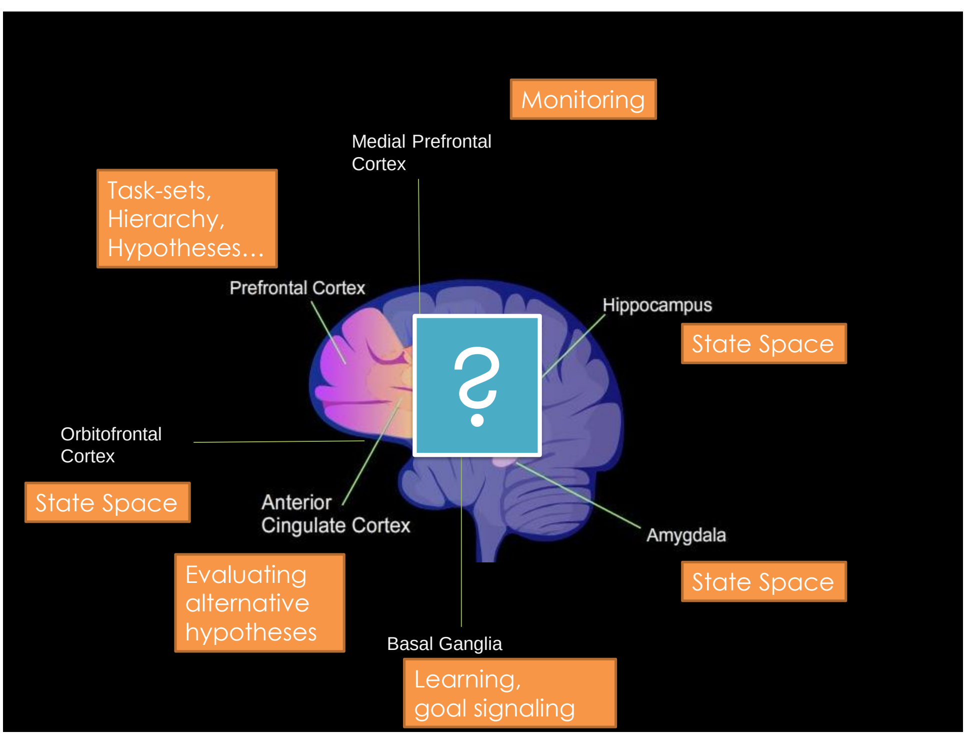
Collins & Frank 2012

Building blocks

1. Structuring the inputs : state spaces
2. Structuring the outputs: action spaces
3. Structuring policies: hierarchy
4. Structuring learning: learning to learn

5. How does the brain do it?

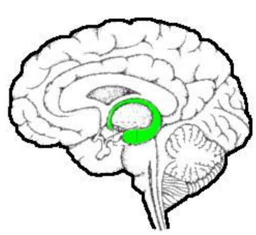
STRUCTURE LEARNING IN THE BRAIN



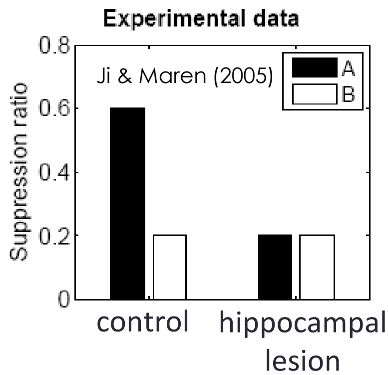
STATE SPACE REPRESENTATION

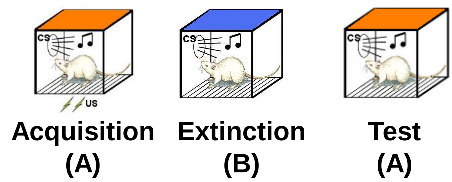
Hippocampus

Necessary for structure learning of latent states

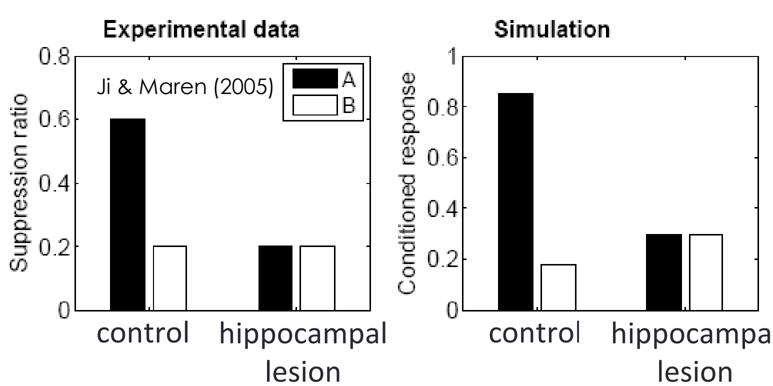


Pre-training lesions of hippocampus abolish renewal





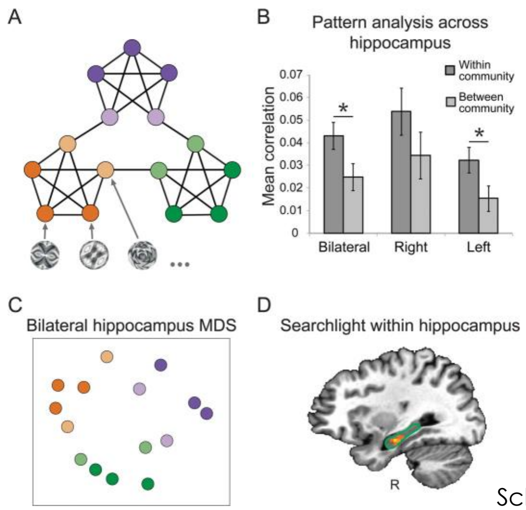
Pre-training lesions of hippocampus abolish renewal



Hippocampal lesions handicap the model's ability to infer new causes

Gershman, Blei & Niv (2010), *Psych Review*

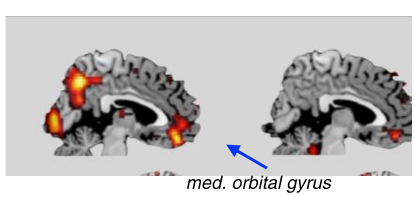
Hippocampus represents temporal community structure – latent state space



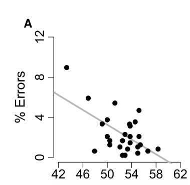
Schapiro et al 2015

OFC

All task relevant information decoded only in OFC



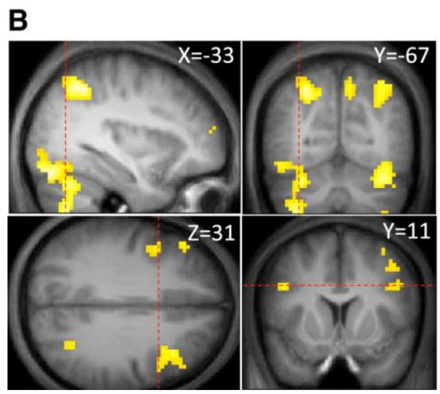
decoding accuracy relates to performance



- Wilson, et al, Neuron, 2014 etc.
- Schuck et al, Neuron, 2016

Fronto-parietal attentional network

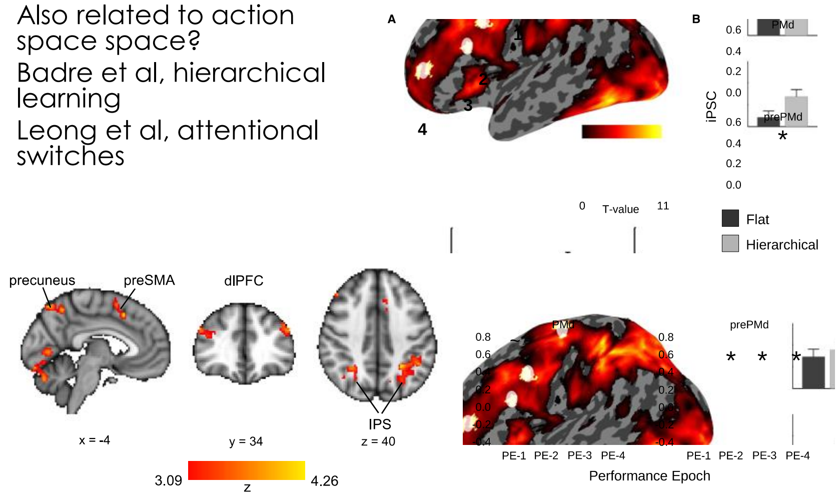
- Niv et al 2015: degree of state space learning correlates with activity in fronto-parietal network



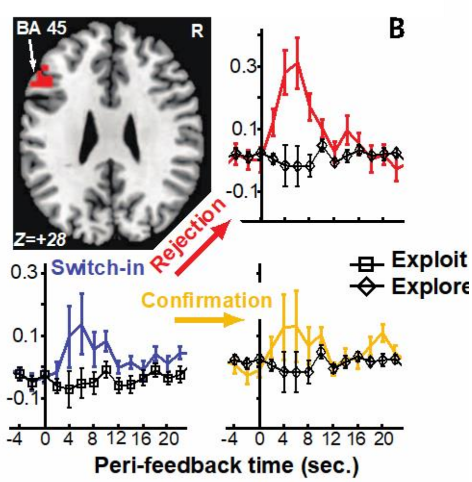
ACTION SPACE REPRESENTATION

Fronto-parietal network

- Also related to action space space?
- Badre et al, hierarchical learning
- Leong et al, attentional switches



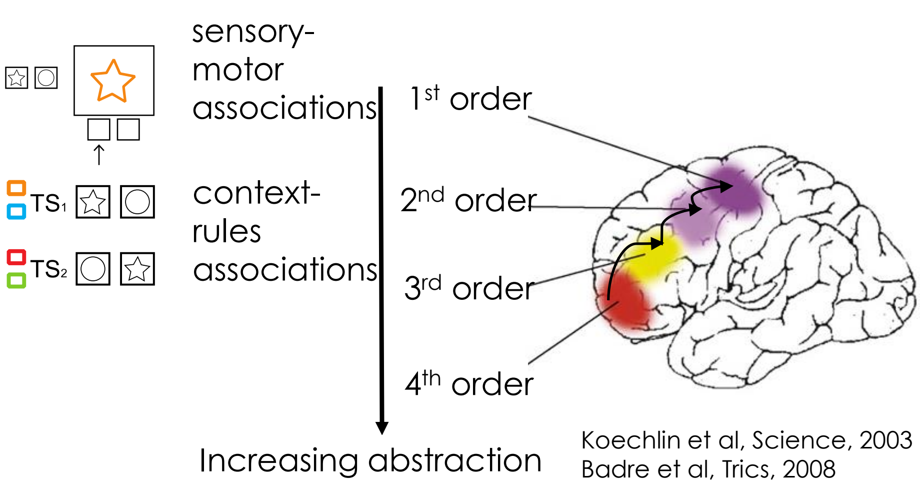
Task-set selection



Donoso et al, 2014

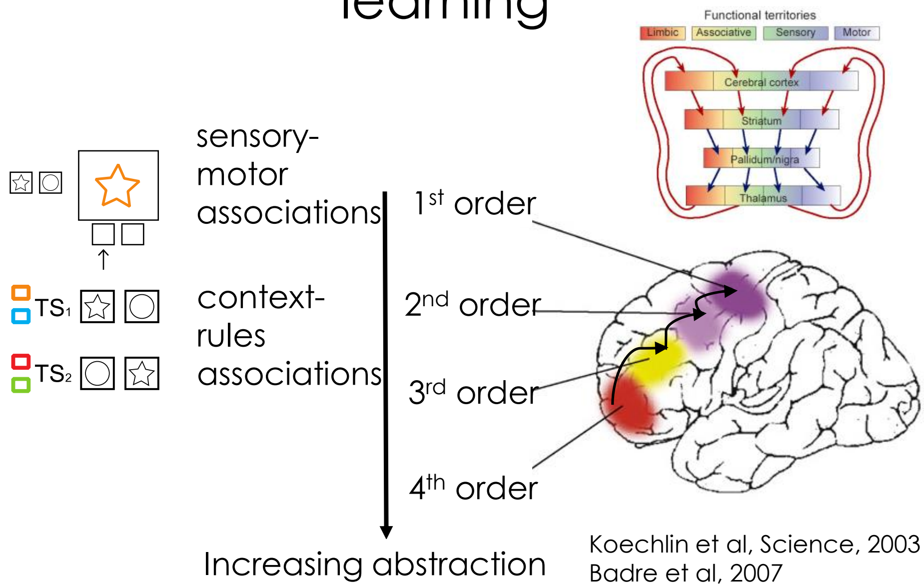
HIERARCHICAL STRUCTURE LEARNING

Hierarchy, rules and abstraction in prefrontal cortex



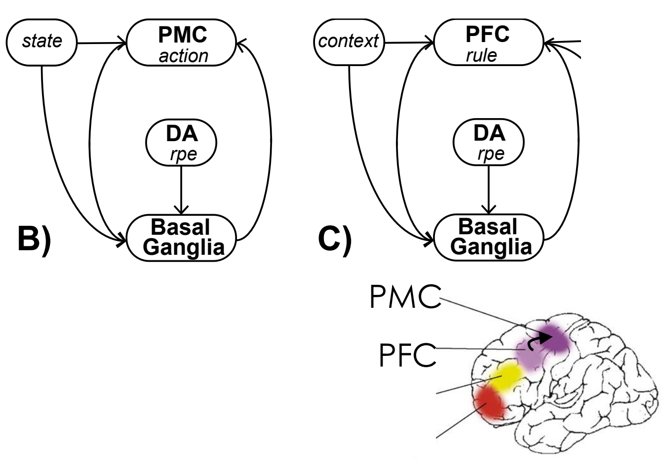
Koechlin et al, Science, 2003
Badre et al, Trics, 2008

Hierarchical reinforcement learning



Koechlin et al, Science, 2003
Badre et al, 2007

Hierarchy in PFC – BG loops



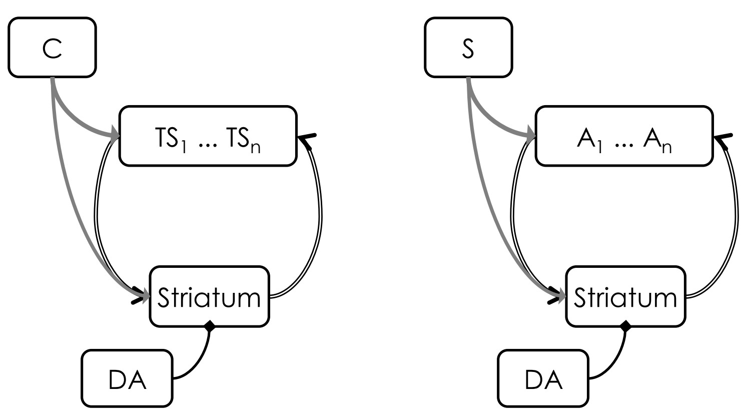
Hierarchy in RL: learning over multiple state/action spaces

$$V(C,TS) \rightarrow \pi(c,TS) = p(TS \mid c)$$

$$V_{TS}(s,a) \rightarrow \pi_{TS}(s,a) = p(a \mid s)$$

$$V(s,a) \rightarrow \pi(s,a) = p(a \mid s,TS)$$

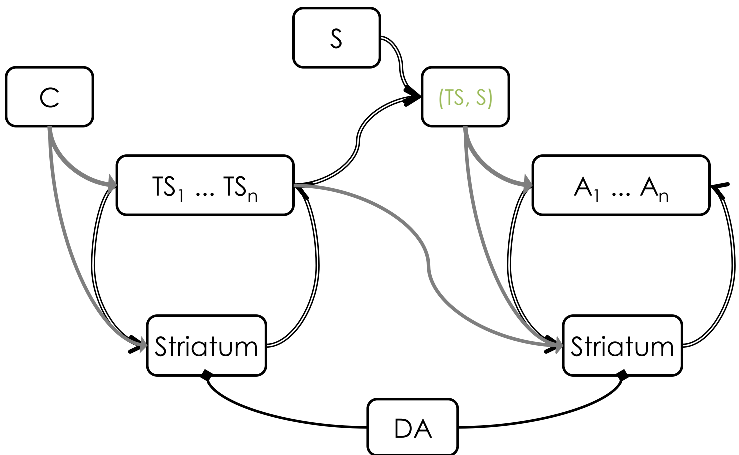
Stimulus-action learning
Context-TS learning



S: stimulus (shape) A: action
BG: basal ganglia DA: Dopaminergic neurons

Collins & Frank 2013, 2016,
Collins, 2017

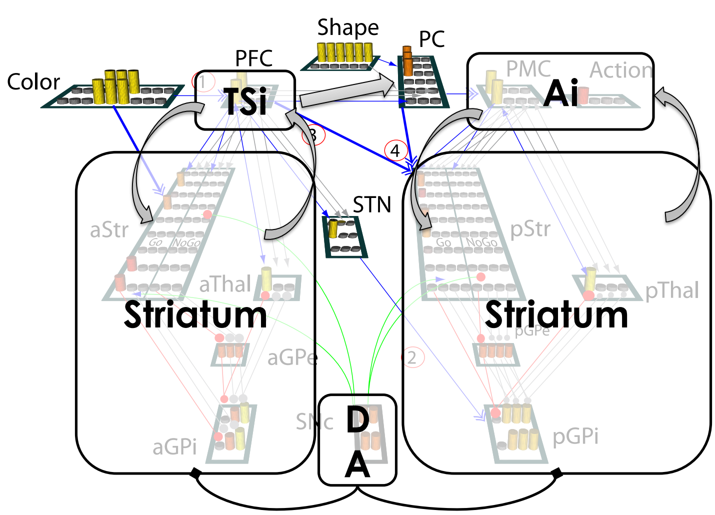
Hierarchical learning network



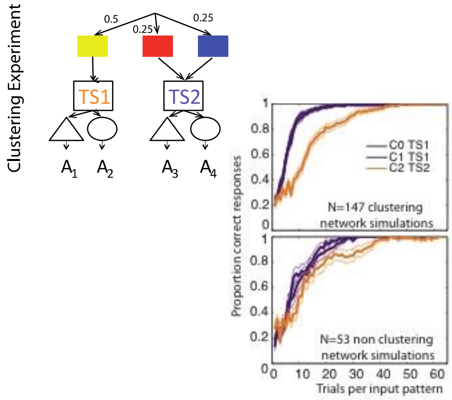
C: context (color) S: stimulus (shape)
TS: task-set A: action
BG: basal ganglia DA: Dopaminergic neurons

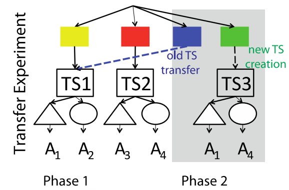
Collins & Frank, Psych Review, 2013

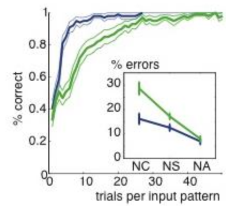
Neurobiologically plausible implementation



Neural Network – Results







Test phase
new contexts:
Old TS
New TS

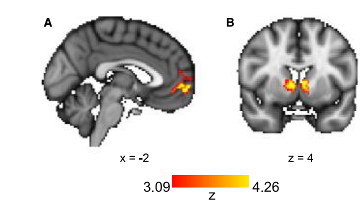
The network learns efficiently unsupervised
Clusters Contexts
Predicts positive, negative transfer

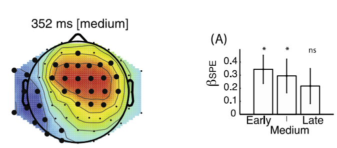
Collins & Frank, 2013

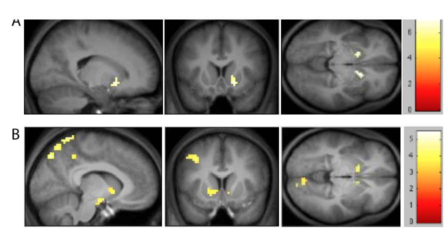
RL network takes structure into account

- EEG (Collins & Frank 2016) and fMRI (Leong et al) RPE signal better explained by structure learning than normal RL

- Evidence for pseudo-reward prediction errors signals (Diuk et al, Ribas-Fernandez et al.)



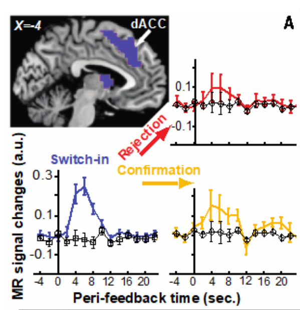


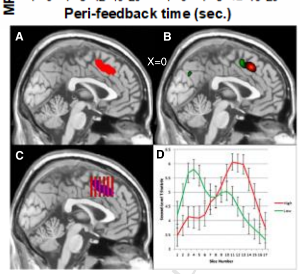


Medial prefrontal cortex also plays an important role in structure learning

- Need for control, monitoring (Donoso et al)
  - task-set reliability
  - task-set creation

- Hierarchical error representation
  - Zarr & Brown, Alexander & Brown





Structure learning in the brain

- Executive network is crucial
- A potential mechanism is multiple parallel reinforcement learning networks, with different state/action representations
- We don't really know how it all fits together… Work on it!

Summary

- We learn both the structure of the environment and the structure of our interactions with the environment
- Structure learning is short-term costly but long-term efficient with generalization, transfer and exploration gains. It occurs by default.
- We don't understand well how the brain supports structure learning.