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
- $P(s' | a, s)$ - Transitions
- $\alpha$ - Starting State Distribution
- $\gamma$ - Discount Factor
- $r(s)$ - Reward [or $r(s, a)$]

Policy $\pi = P(\alpha | s)$

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) \( \mathcal{A} = \{ \text{“pick star”, “pick circle”} \} \)
   \( \{ \text{“pick blue”, “pick red”, …} \} \)

Wilson & Niv 2011, …

With what algorithm should I learn?

Learning algorithm
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?

- Real life learning suffers from the curse of dimensionality
- Structure learning: Compressing the environment into a small, relevant state space

Hypothesis: 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?

“It’s that time of year when guys randomly explode.”

Slide from S Gershman
Some possibilities

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.
Conditioning as clustering

Case study: renewal

Acquisition (box A) Extinction (box B)
Conditioned responding is renewed!

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

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**.
Experimental design

Conditioning (3 trials) → 24 hours → Extinction (24 trials) → 24 hours → Long-term memory test (4 trials) → 30 days → Spontaneous recovery (4 trials)

Slide from S Gershman

Conditioning (3 trials) → 24 hours → Extinction (24 trials) → 24 hours → Long-term memory test (4 trials) → 30 days → Spontaneous recovery (4 trials)

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

Bottlenecks as optimal subgoals for hierarchical structure

Schapiro et al 2013

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 levels

Hierarchical reinforcement learning
HIERARCHY
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

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

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

**learning task-sets**

- C1, C2, C3, C4, C5, C6
- 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

Creating abstract action spaces, latent context
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

The actor

Task-sets stored in long-term memory

Monitoring buffer

Behavior
Exploitation

The actor

Reliability

Task-sets stored in long-term memory

Monitoring buffer

Behavior

Exploitation

The actor

Reliability

Task-sets stored in long-term memory

Monitoring buffer

Less recently used

Behavior
Switch from Exploitation to Exploration
Return to Exploitation (Rejection events)

The actor

Reliability

The probed actor is rejected!

Monitoring buffer

Task-sets stored in long-term memory

The probe actor is confirmed!

Monitoring buffer

Discarded from the monitoring buffer

Behavior
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

Latent task-set space

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 overridden by strong prior that structure learning is long-term beneficial

Collins, JoCN, 2017
Phase 1

- $A_1$ (correct)
- $A_2$
- $A_1$
- $A_3$
- $A_4$

Phase 2

- $A_3$ (incorrect)
- $A_3$
- $A_2$

P = 1/4

P = 1/4

P > 1/4

Model simulations

Subjects (N=34)

Fit Parameters

- model CSTS2
- Block 1 only
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 $\rightarrow$ TS1 than for C3 $\rightarrow$ 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

A) B) C)

\[
\begin{align*}
  s & \rightarrow \text{Learning algorithm} \\
  \text{Learning algorithm} & \rightarrow r \\
  r & \rightarrow a \\
  a & \rightarrow s
\end{align*}
\]
Learning the structure of a task

(a) 11
   A G O Π

(b) Context First (CF)

11 + A + Π
   A O G Π

(c) Group 1: CF -> CF
Group 2: CL -> CF
Group 3: CL -> CL
Group 4: CF -> CL

HARRY F. HARLOW

Monkeys and children

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 $\rightarrow$ reduce curse of dimensionality
- Generalize $\rightarrow$ learn more flexibly
- Explore $\rightarrow$ learn more faster
- Adjust $\rightarrow$ learn more efficiently

Simplifying the problem

Badre et al 2010, Frank & Badre 2011
Abstract rules: transfer and generalization

Collins & Frank 2012, 2016

Rule 1

Rule 2

Transfer

Generalization

Subjects (N=34)

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

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

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
Hierarchy, rules and abstraction in prefrontal cortex

Koechlin et al, Science, 2003
Badre et al, Trics, 2008
Hierarchical reinforcement learning

- 1st order
- 2nd order
- 3rd order
- 4th order

Increasing abstraction

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

Hierarchy in PFC – BG loops
Hierarchy in RL: learning over multiple state/action spaces

\[ V(C, TS) \rightarrow \pi(c, TS) = p(TS | c) \]

\[ \pi_{TS}(s, a) \rightarrow \pi_{TS}(s, a) = p(a | s)_{TS} \]

Stimulus-action learning
Context-TS learning

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

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.