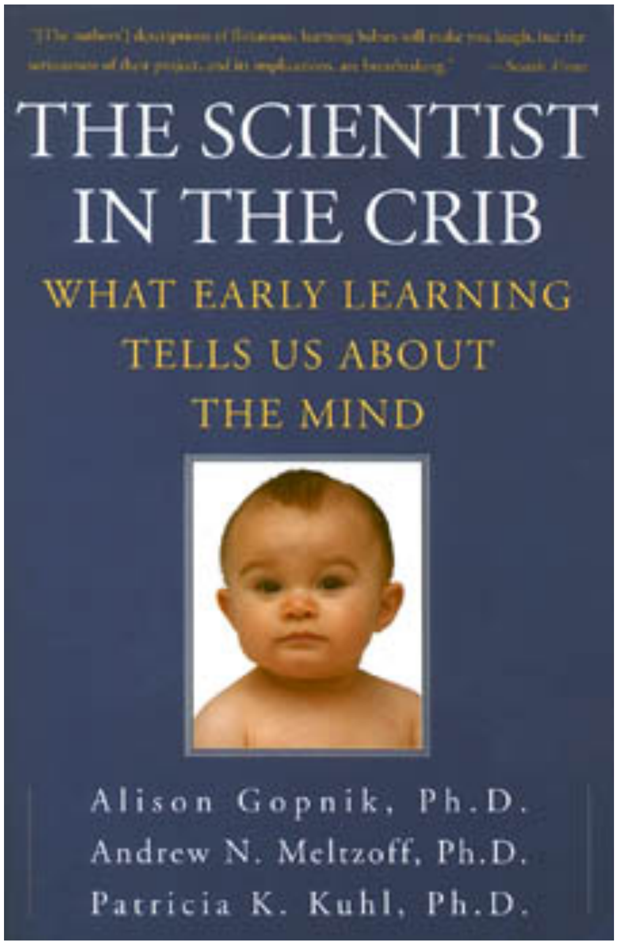
Information-theory based policies for learning in closed sensori-motor loops

Friedrich T. Sommer
UC Berkeley - Redwood Center for Theoretical Neuroscience

Lecture 2022

WHAT MAKES US EXPLORE AND PLAY?



Talk Outline and Collaborators

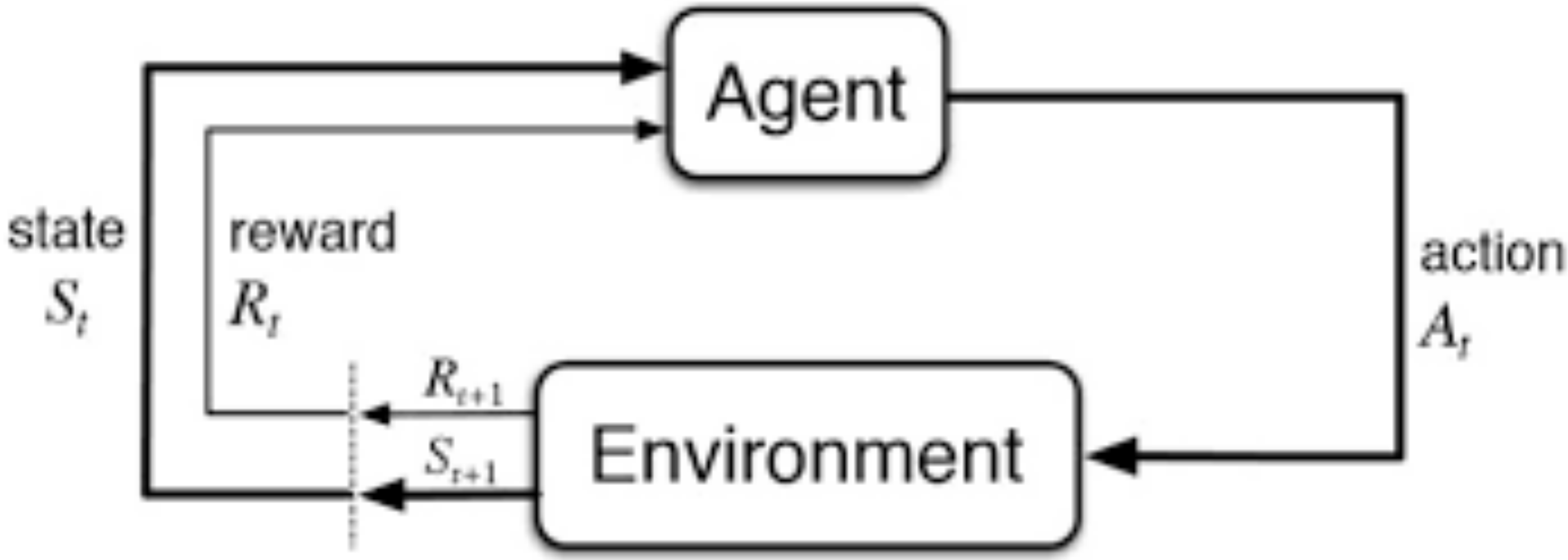
1) Intro: Theories on information-driven exploration

2) Exploration based on predicted information gain (PIG)

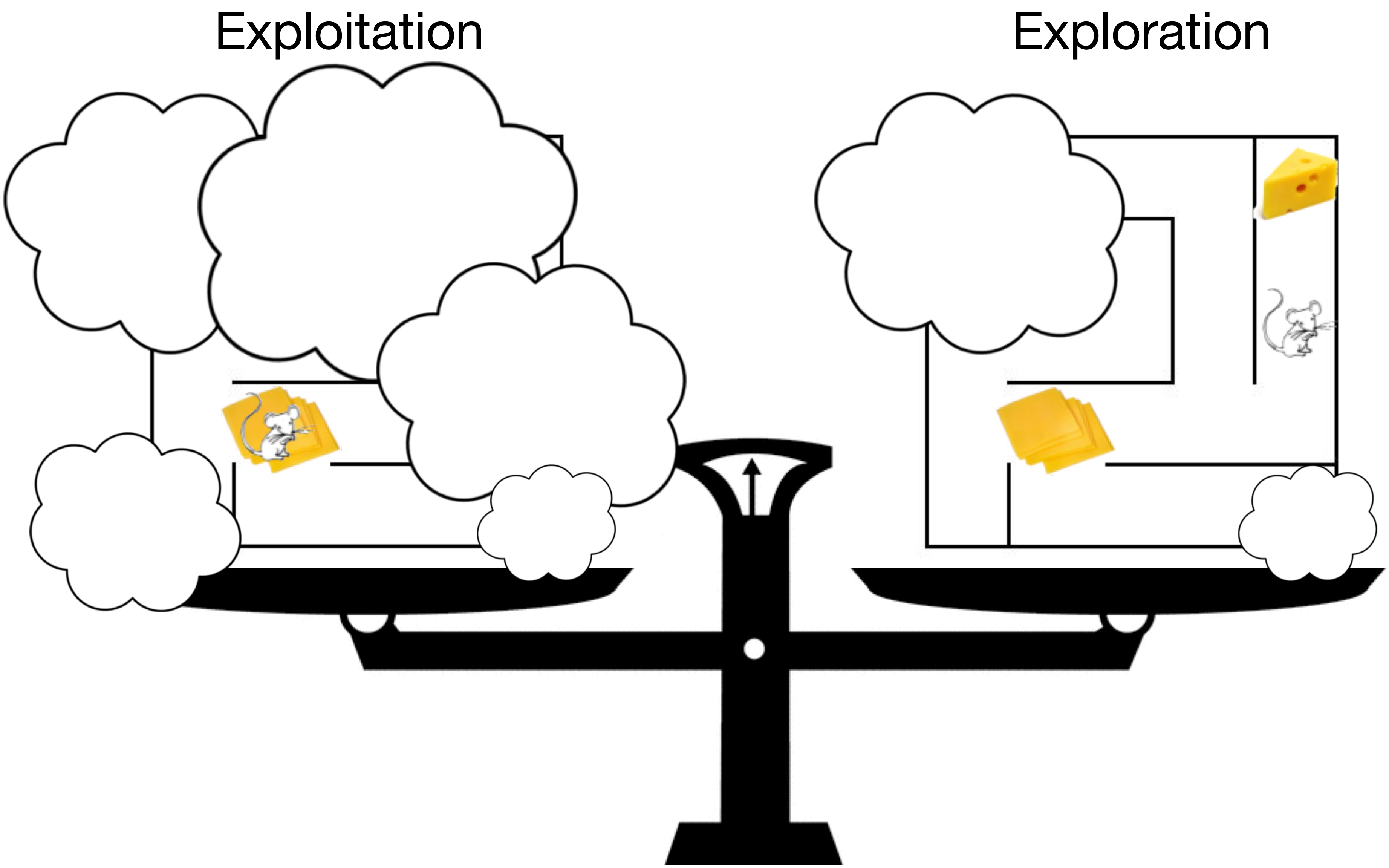
3) PIG exploration in unbounded state spaces

Daniel Little   Shariq Mobin   James Arnemann

WHY DO WE EXPLORE?
REINFORCEMENT LEARNING PERSPECTIVE

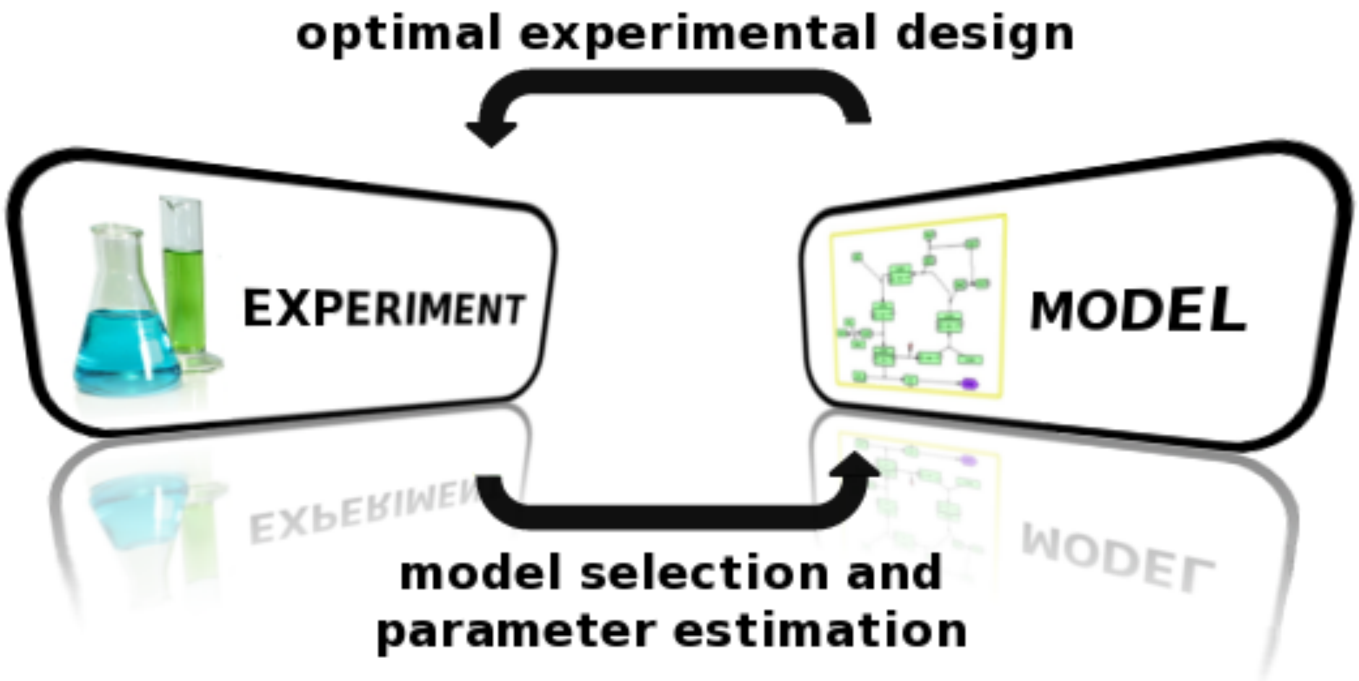


WHY DO WE EXPLORE?
REINFORCEMENT LEARNING PERSPECTIVE



WHY DO WE EXPLORE?

PERSPECTIVE OF OPTIMAL DESIGN – ACTIVE LEARNING



- Theory for learning in the brain (from little data)
- Basis for building machines that can autonomously learn
- Balance: "modeling the predictable" versus "discovering the novel"

TIME LINE OF EARLY THEORETICAL WORK

Kirstine Smith 1918: Optimal experimental design

On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations (Biometrika 1918)

D.V. Lindley 1956: Bayesian definition of information gain

On a measure of information provided by an experiment (Annals of Math. Stat. 1956)

E. Pfaffelhuber 1972: Definition of missing information

Learning and information theory (Intern, J. Neurosci. 1972)

MacKay 1992:

Information-based objective functions for active data selection (Neural Comp. 1992)

Oaksford & Chater 1994:

Expected information gain in selection tasks

THE BAYESIAN VIEW

Parameterized probabilistic model of observation:

$$p(x) = \int p(x \mid \theta) p(\theta) \, d\theta$$

Information gain by individual observation (Lindley 56):

$$I(x) = \int p(\theta \mid x) \log p(\theta \mid x) \, d\theta - \int p(\theta) \log p(\theta) \, d\theta$$

with prior $p(\theta)$ and posterior $p(\theta \mid x)$ of parameter

.

Average information gain (Lindley 56):

$$I = E_x I(x)$$
$$= \int dx p(x) \int p(\theta \mid x) \log p(\theta \mid x) d\theta - \int p(\theta) \log p(\theta) d\theta$$
$$= KL[p(x,\theta) \parallel p(x) p(\theta)]$$

(Shannon channel capacity)

$$= E_x KL[p(\theta \mid x) \parallel p(\theta)]$$
$$= E_\theta KL[p(x \mid \theta) \parallel p(x)]$$

Note that:

$$I = H(\theta \mid x) - H(\theta) = E_x KL[p(\theta \mid x) \parallel p(\theta)]$$

Information gain for active data selection (MacKay 92):

Sampling observation history: $h_N = \{a_i, s_i\}, i = 1, \ldots, N$

Estimated posterior distribution: $\hat{p}_N(\theta) = p(\theta \mid h_N)$

(Bayesian) information gain:

$$I = E_{s_{N+1}} KL[\hat{p}_{N+1}(\theta) \,||\, \hat{p}_N(\theta)] = E_{s_{N+1}} (H_N - H_{N+1})$$

Past BIG was renamed Bayesian surprise (Baldi & Itti, 2004)

MISSING INFORMATION

Little & Sommer, Frontiers in Neural Circuits 2013 (ArXiv 2011)

Estimate of ground truth probability:

$$\hat{p}(x) \approx p(x)$$

Missing information of current estimate (Pfaffelhuber 72):

$$I_M(\hat{p}) = KL[\,p(x) \,||\, \hat{p}(x)]$$

Information gain:

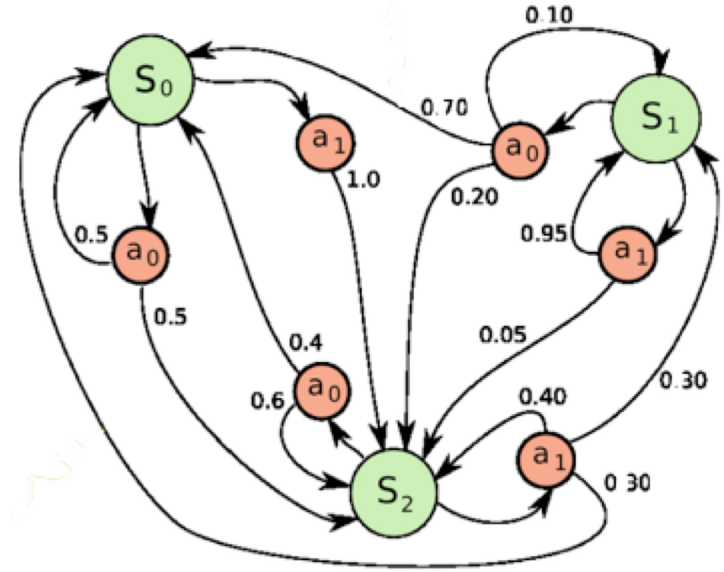
$$I = I_M(\hat{p}_N) - I_M(\hat{p}_{N+1})$$

Past IG "curiosity" based policies in agents (Storck et al. 1995)
Predictive IG = PIG (Little & Sommer 2011, 2013)

MODEL OF THE ENVIRONMENT

Controllable Markov Chain

$$\{A, S, p(s' \mid a, s)\}$$



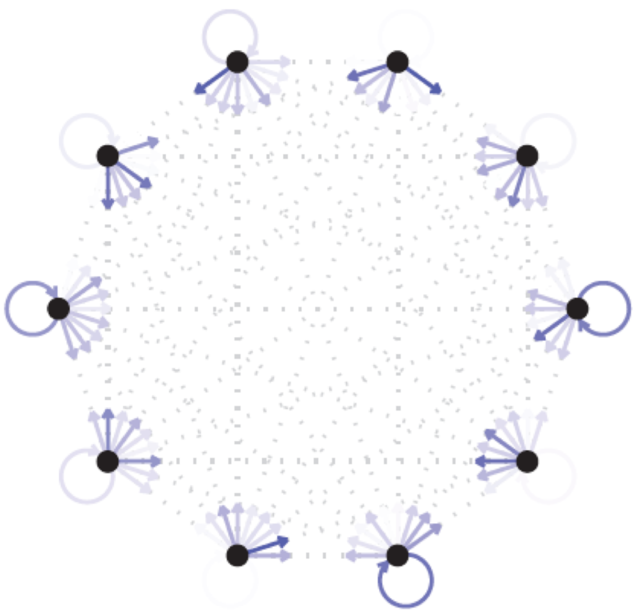
Learning task:

Estimate: $\hat{p}(s' \mid a, s) = \hat{\Theta}_{a,s,s'} \approx p(s' \mid a, s) = \Theta_{a,s,s'}$

THREE TEST ENVIRONMENTS

LEARNING ACROSS A RANGE OF STRUCTURES

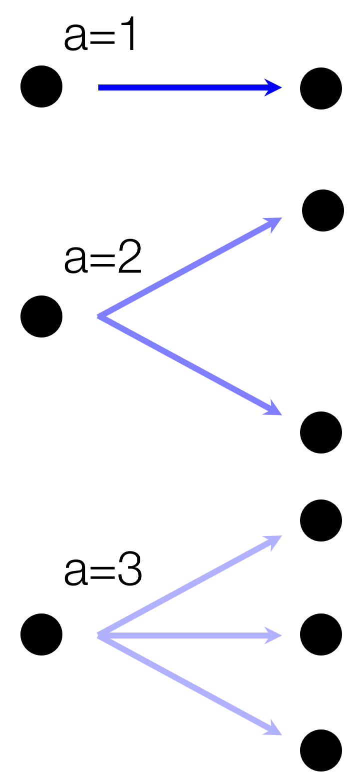
Dense Worlds



ex. a=1

Mazes

1-2-3 Worlds



$$f(\Theta_{a,s,\cdot}) = Dir(\alpha) = \frac{1}{Z(\alpha)} \prod_{s'} (\Theta_{a,s,s'})^{\alpha_{s'}-1}$$

LEARNING-DRIVEN EXPLORATION

TWO SEPARATE CHALLENGES

1.Inference: Given a set of data, what is the best estimate $\widehat{\Theta}$ of Θ

2. Exploration: Given $\widehat{\Theta}$, how should an agent choose actions to best improve the estimate

1. INFERENCE

Missing information for entire CMC:

$$I_M = \sum_{a,s} \alpha_{as} KL[\Theta_{a,s,s'} \,||\, \hat{\Theta}_{a,s,s'}]$$

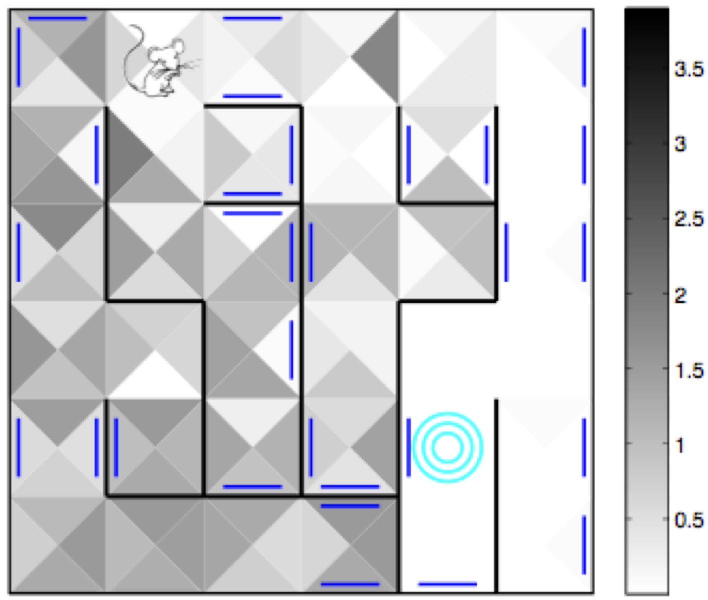
Theorem for inference step:
Bayesian inference minimizes missing information:

$$\hat{\Theta} = E_{\Theta|h}[\Theta] = \arg\min_{\Phi} E_{\Theta|h}[I_M(\Phi)]$$

LEARNING DURING EXPLORATION

MISSING INFORMATION IS UNEVENLY DISTRIBUTED

Missing Information

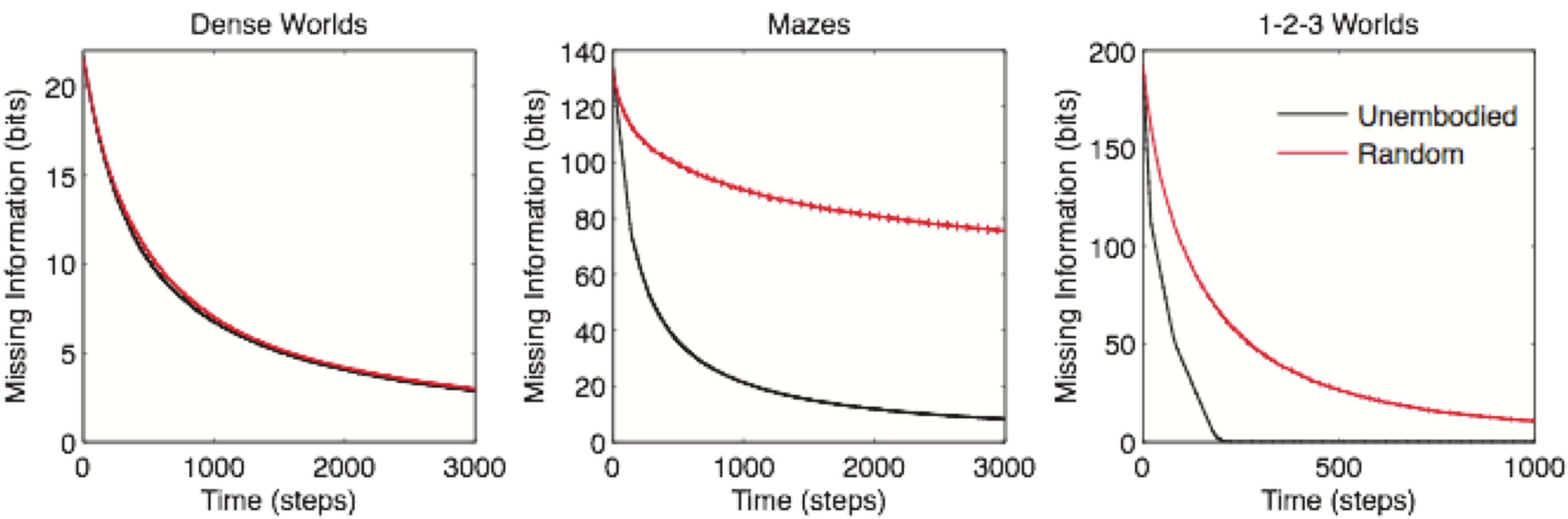


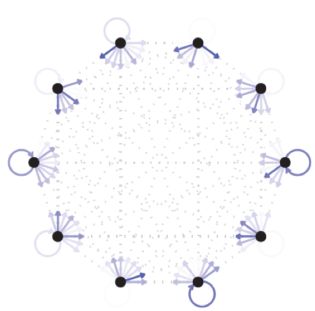
Control Strategies:

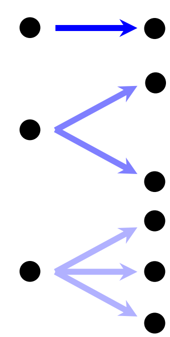
Random Action - an undirected baseline learner

Unembodied - an upper bound on learning

COMPARING CONTROLS
THE EMBODIMENT CONSTRAINTS ON LEARNING







2. OBJECTIVE FOR EXPLORATION

PREDICTED INFORMATION GAIN

1. Use current model to predict next sensory input: $\hat{\Theta}_{a,s,s'}$

2. Add fictive new observation s' to current model: $\hat{\Theta}^{a,s \longrightarrow s'}$

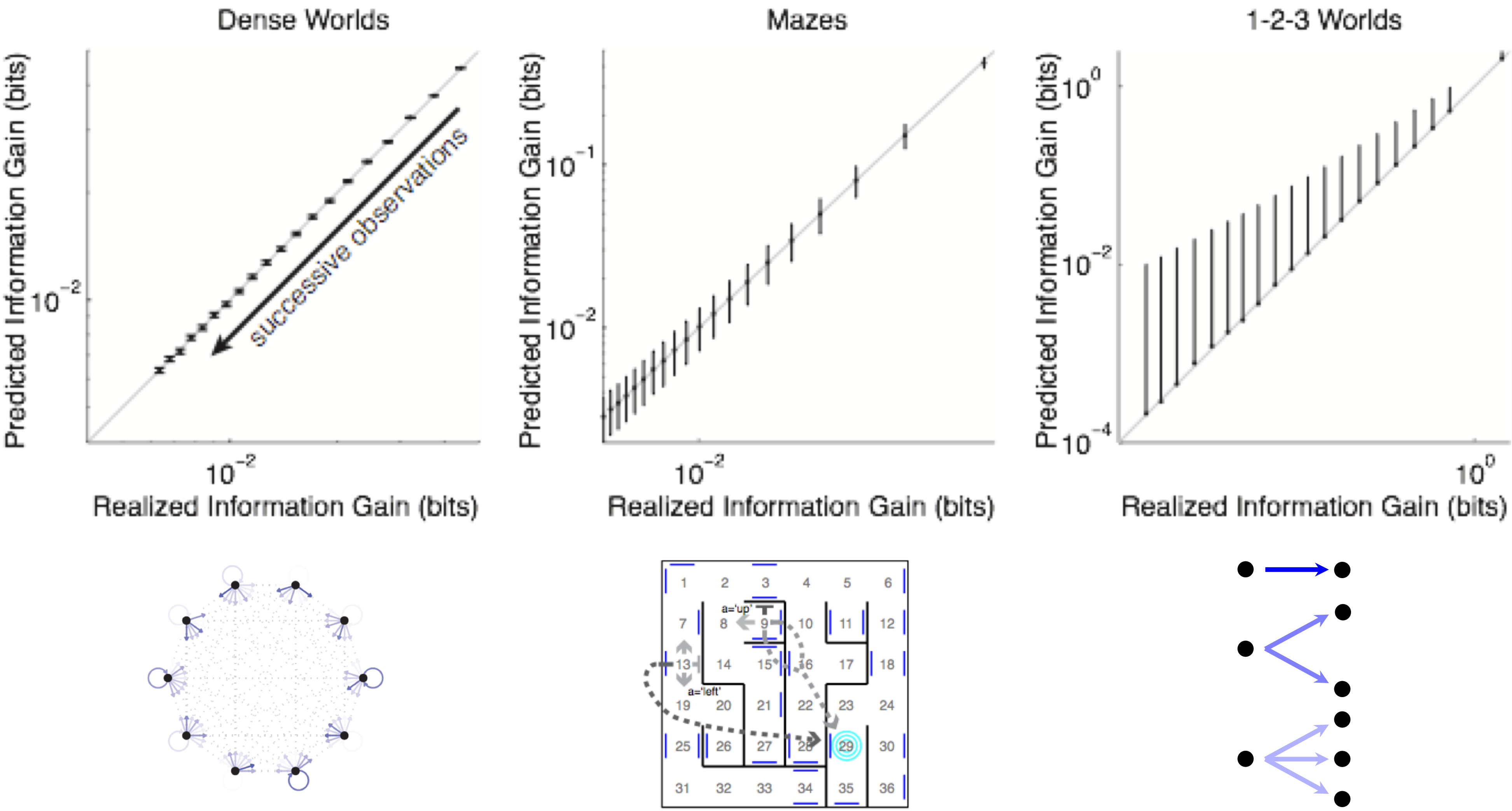
3. Compute predicted information gain:

$$PIG(a,s) = E_{s',\Theta|h}[KL[\Theta \,||\, \hat{\Theta}] - KL[\Theta \,||\, \hat{\Theta}^{a,s \longrightarrow s'}]]$$

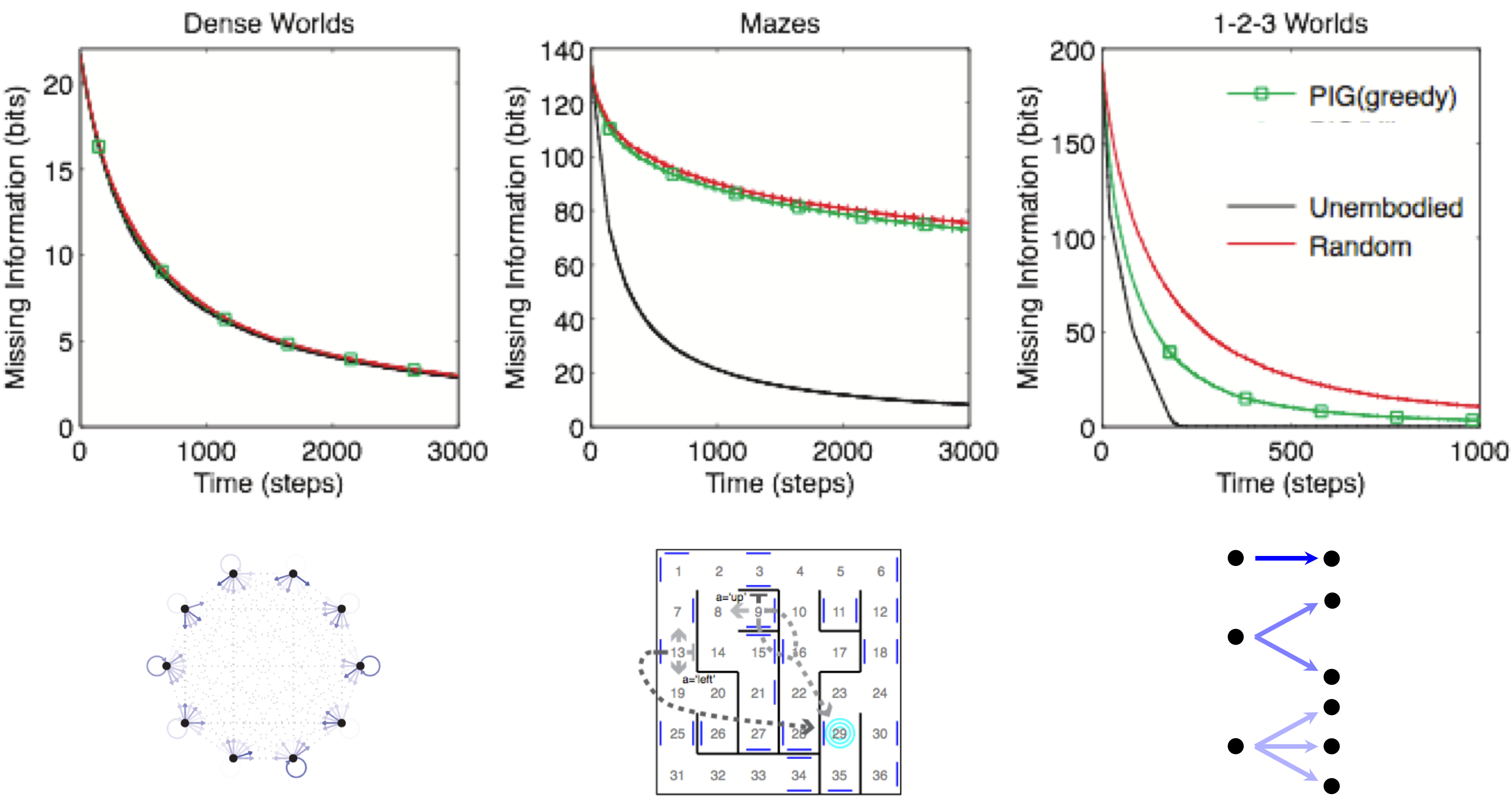
$$= \sum_{s'} \hat{\Theta}_{a,s,s'} KL[\hat{\Theta}^{a,s \longrightarrow s'} \,||\, \hat{\Theta}]$$

PREDICTED INFORMATION GAIN (PIG)

ACCURATE ESTIMATION OF LEARNING VALUE



GREEDY MAXIMIZATION OF PIG
IMPROVES LEARNING IN 1-2-3 WORLDS



OPTIMIZATION WITH LONGER TIME HORIZON

Forward search for optimal policy has exponential complexity

Use value iteration (Bellman 1957)

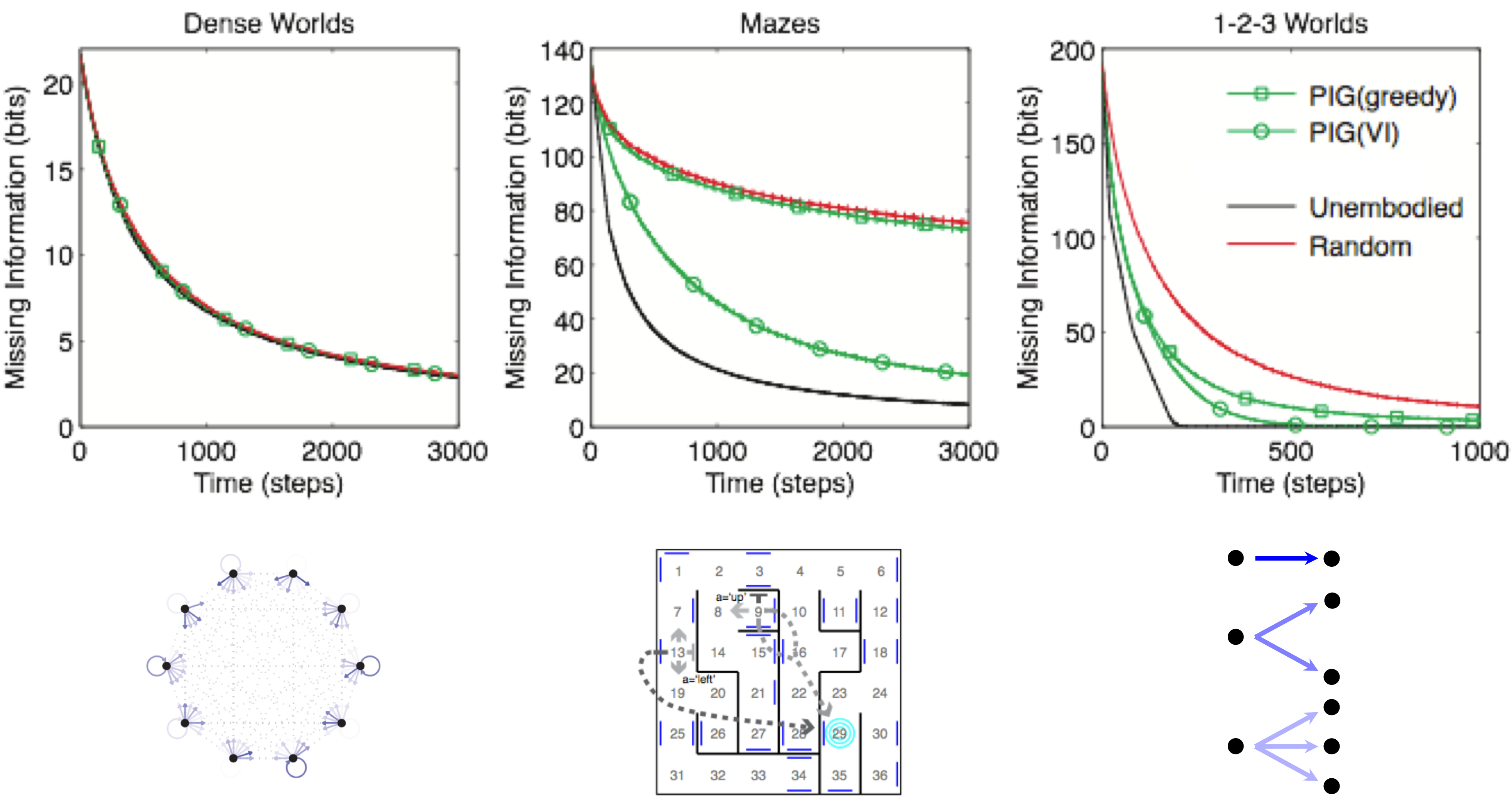
$$Q_0(a,s) = PIG(a,s)$$

$$Q_{\tau-1}(a,s) = PIG(a,s) + \eta \sum_{s' \in S} \hat{\Theta}_{ass'} V_\tau(s')$$

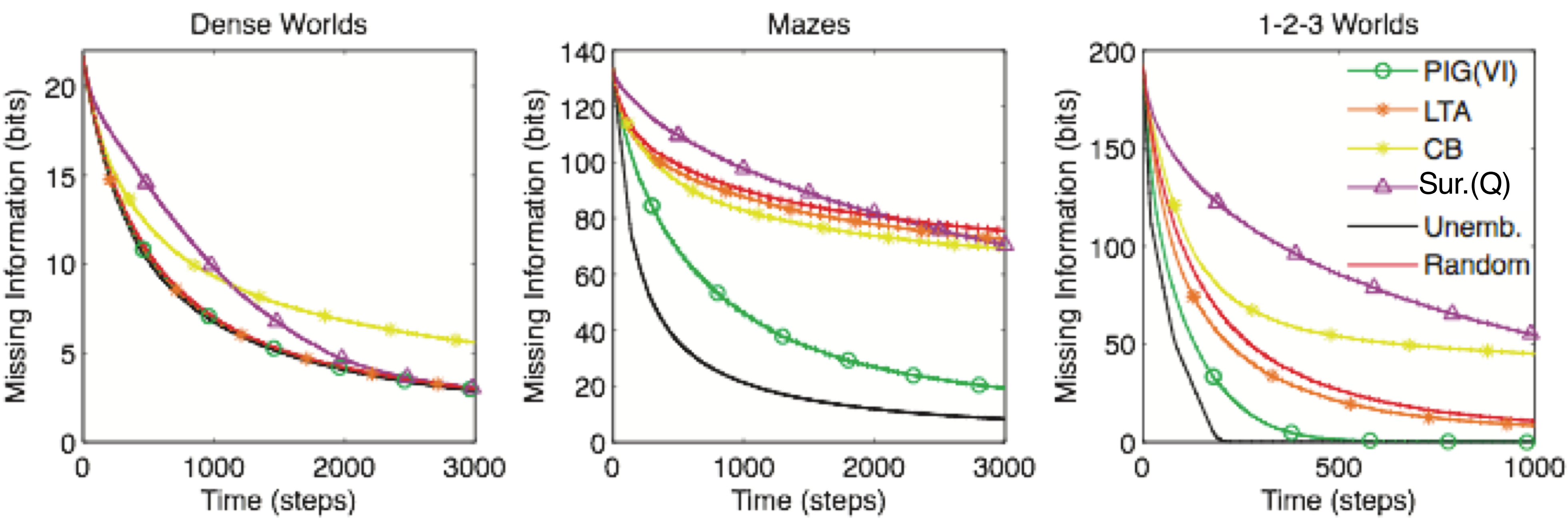
$$V_\tau(s') = \max{}_a \, Q_\tau(a,s)$$

VALUE ITERATED MAXIMIZATION OF PIG

CLOSING THE EMBODIMENT GAP



PREVIOUS EXPLORATION STRATEGIES
PIG(VI) OUTPERFORMS ALTERNATIVES IN STRUCTURED WORLDS



Previous Strategies:

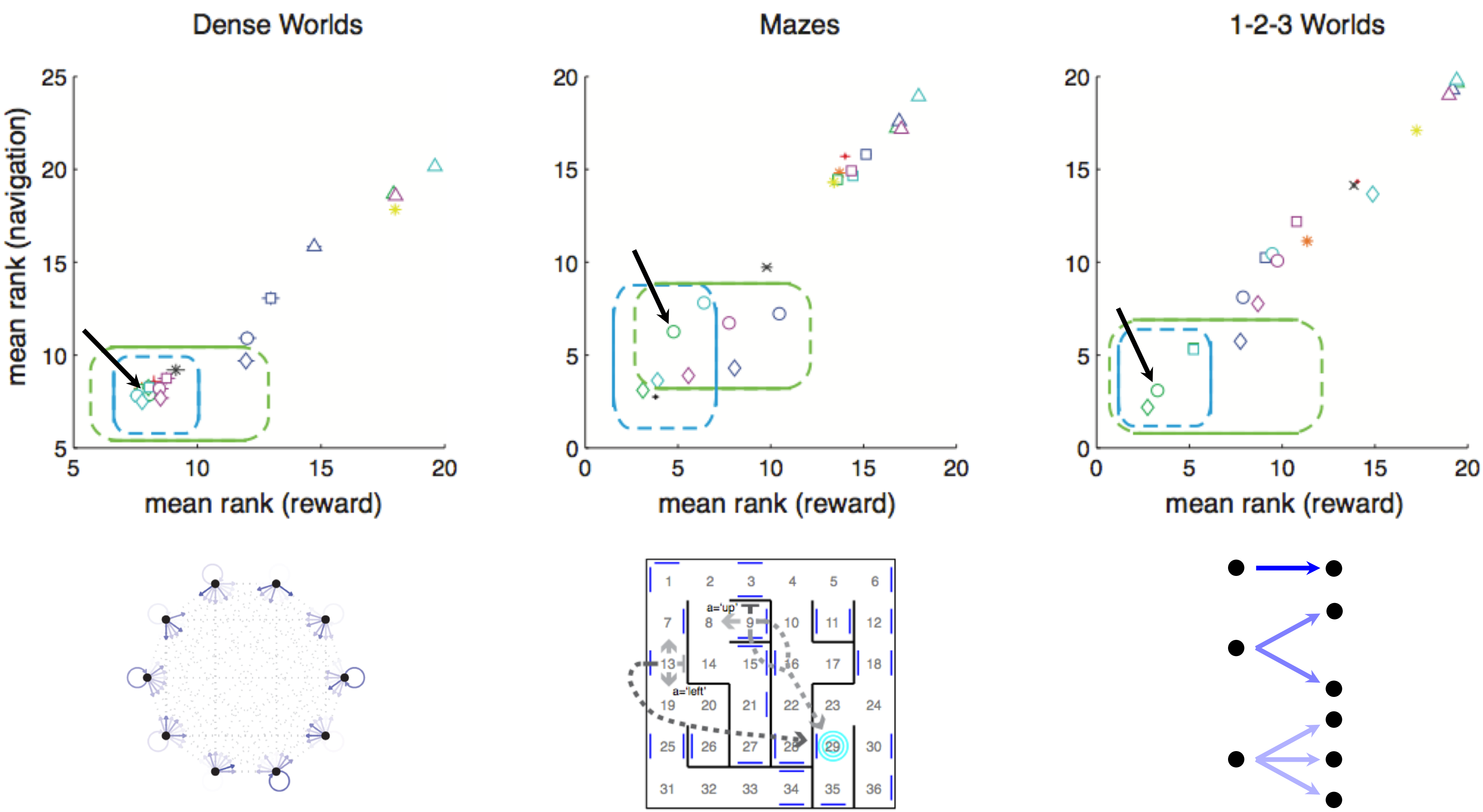
Least Taken Action (LTA) - Si, Herrmann, and Pawelzik 2007

Counter Based (CB) - Thrun 1992

Q-Learning on Past Change (PC(Q)) - Storck, Hochreiter, Schmidhuber 1995

GENERAL UTILITY OF EFFICIENT LEARNING

INDEPENDENT GOAL-DIRECTED TASKS



PIG IN UNBOUNDED STATE SPACE

Mobin, Arnemann, Sommer, NIPS 2014

Chinese Restaurant Process – CRP-PIG

$$p_i(C_t) = \frac{c_i}{t+\theta}, i = 1,\dots,K_t$$

$$p_\psi(C_t) = \frac{\theta}{t+\theta}$$

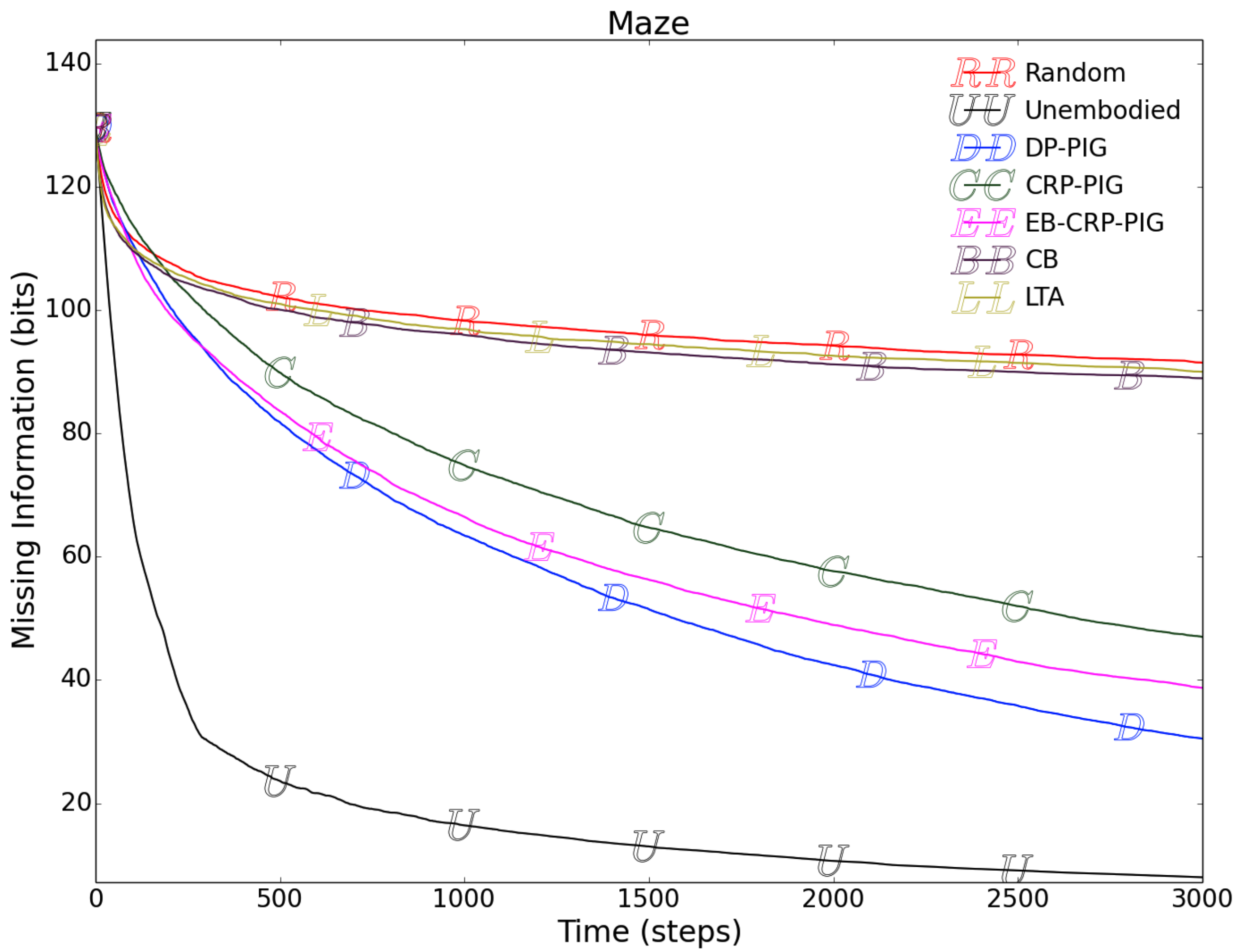
Empiricial Bayes Version – EB-CRP-PIG

$$\theta(t) \approx \frac{K_t}{\ln(t) + \gamma + \frac{1}{2t} - \frac{1}{12t^2}}$$

with $\gamma \approx 0.577$ Euler's Mascheroni constant

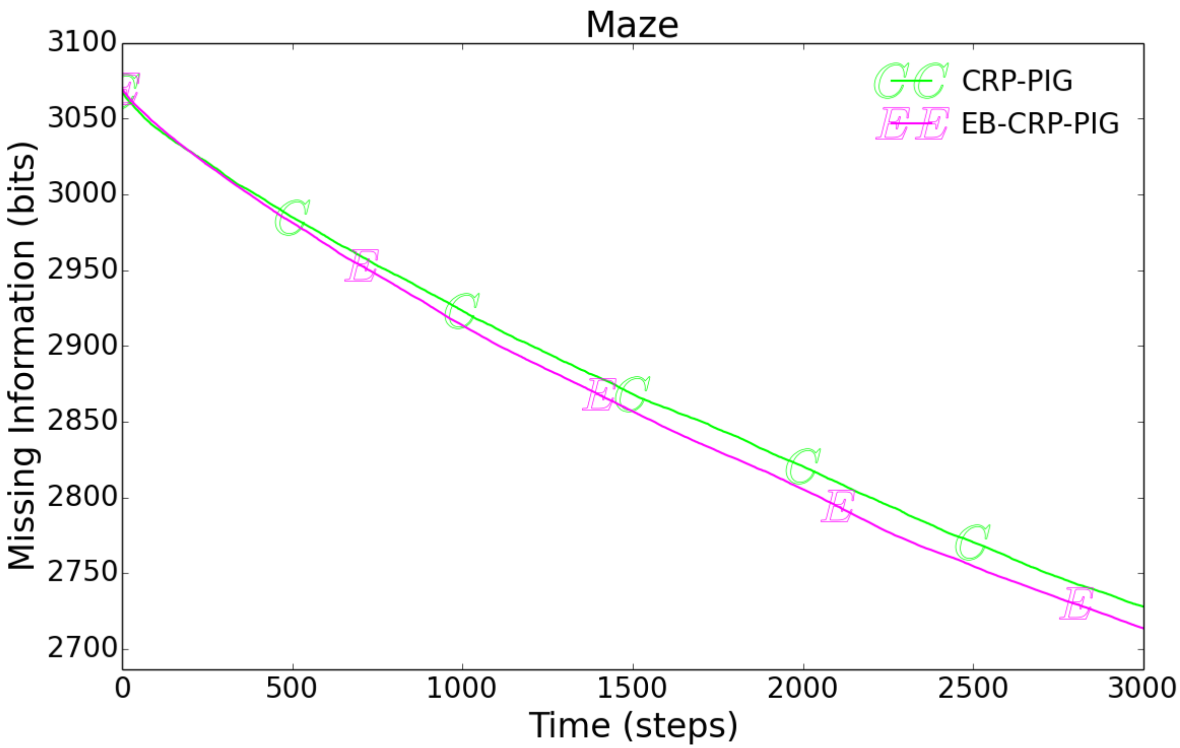
PIG IN UNBOUNDED STATE SPACE

Mobin, Arnemann, Sommer, NIPS 2014



Reduction in Missing Information in bounded maze

PIG IN UNBOUNDED STATE SPACE
Mobin, Arnemann, Sommer, NIPS 2014



Reduction in Missing Information in unbounded maze

PIG IN UNBOUNDED STATE SPACE

Exploration in unbounded environment

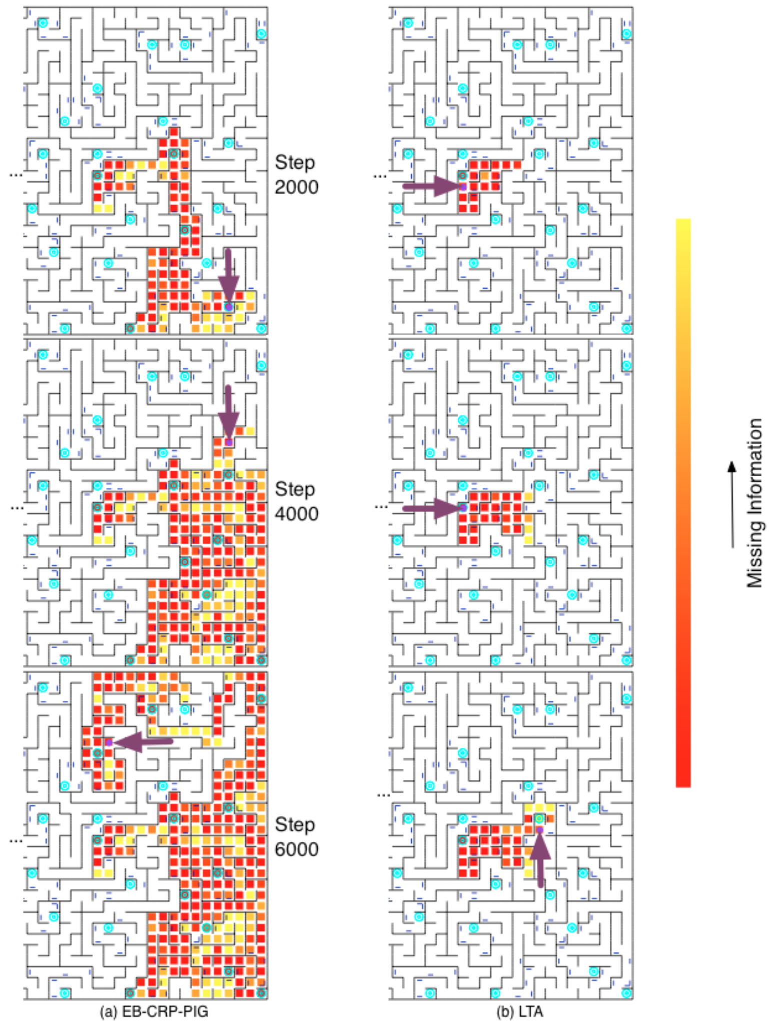


Figure 6: Unbounded Maze environment. Exploration is depicted for two different agents (a) EB-CRP-PIG and (b) LTA, after 2000, 4000, and 6000 exploration steps respectively. Initially all states are white (not depicted), which represent unexplored states. Transporters (blue lines) move the agent to the closest gravity well (small blue concentric rings). The current position of the agent is indicated by the purple arrow.

SUMMARY

1. Learning in action-perception loops is the old optimal design problem

2. State-space information gain (PIG) versus Bayesian information gain

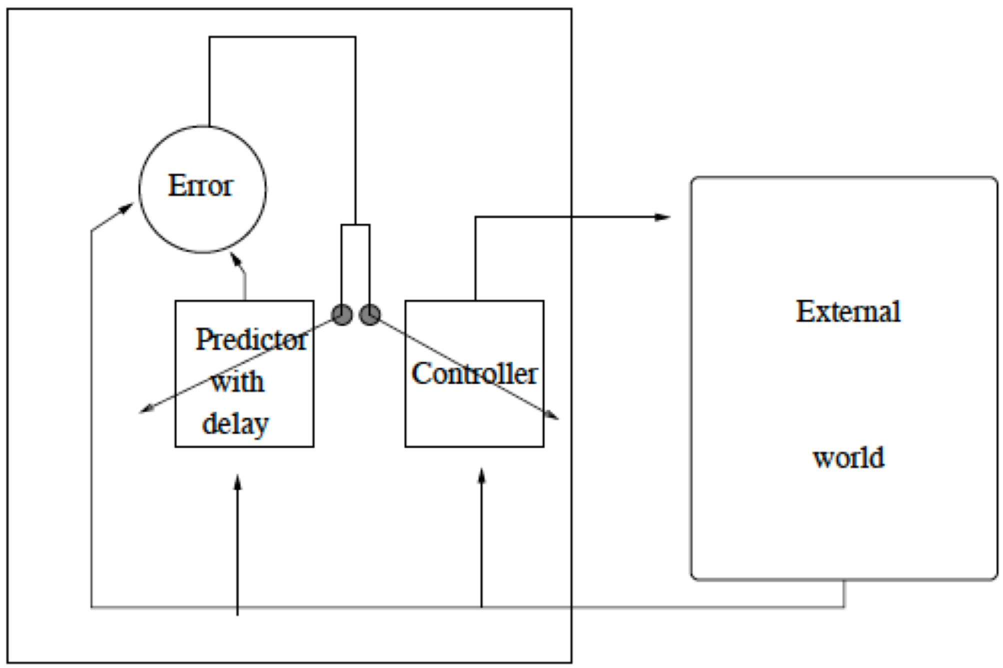
3. Maximizing PIG minimizes missing information

4 Nongreedy optimization is critical in interesting environments

5. Extension of PIG to unbounded environments:
    CRP+ Empirical Bayes works best
    Surprise-based information seeking does not eliminate surprise
    Balance between eliminating uncertainty and discovering more states is model depending

Other objectives for exploration

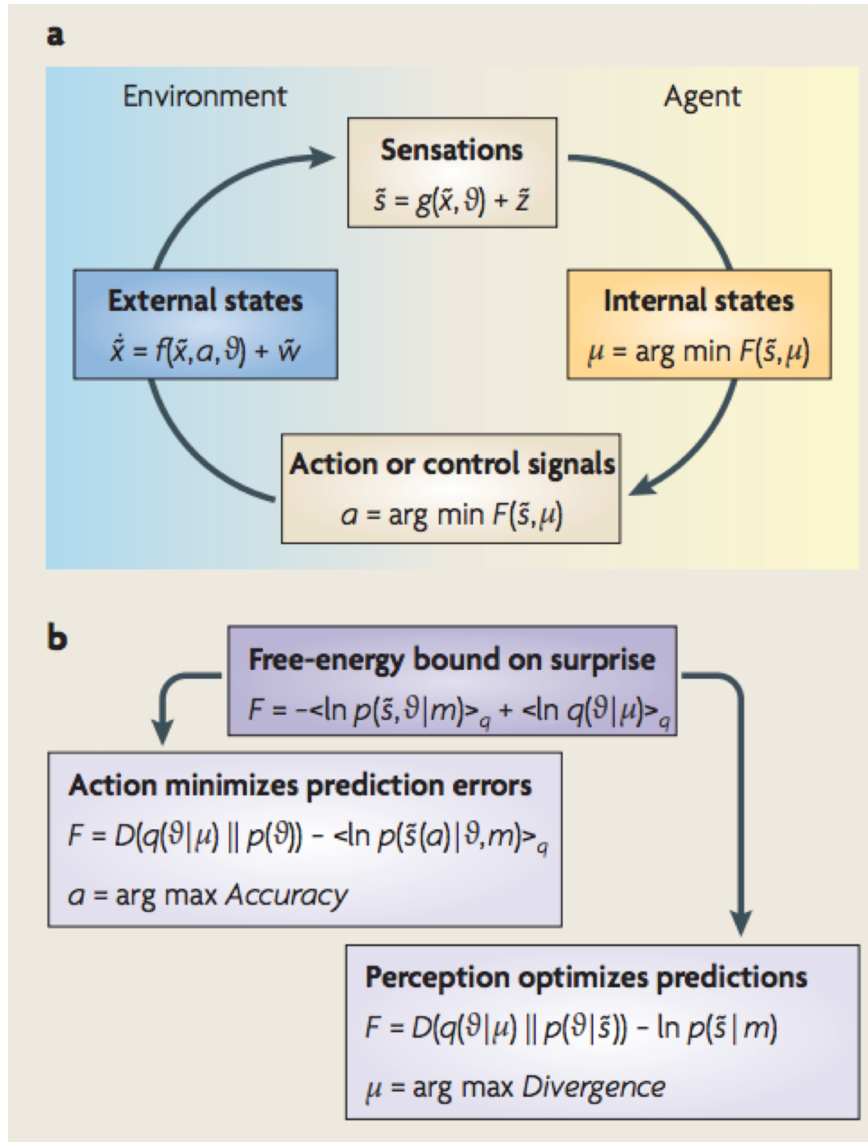
1. Homekinesis (Der, 2000)



Other objectives for exploration

1. Free Energy Principle
(Friston, 2010)

-> Dark corner problem



Other objectives for exploration

Learning Progress (Kaplan & Oudeyer, 2007)

= derivative of prediction error

Does not require information estimation and still avoids problems of policies of directly minimizing prediction error

MEASURES OF UTILITY TOWARDS LEARNING

IN THEORETICAL PSYCHOLOGY

Predicted Information Gain (PIG) (Oaksford & Chater 1994)

$$\sum_{s'} \widehat{\Theta}_{a,s,s'} D_{\mathrm{KL}}(\widehat{\Theta}^{a,s\to s'} \,||\, \widehat{\Theta})$$

Predicted Mode Change (PMC) (Baron 2005, Nelson 2005)

$$\sum_{s'} \widehat{\Theta}_{a,s,s'} \left[\max_{s^*} \widehat{\Theta}^{a,s\to s'}_{a,s,s^*} - \max_{s^*} \widehat{\Theta}_{a,s,s^*}\right]$$

Predicted L1 Change (PLC) (Klayman & Ha 1987)

$$\sum_{s'} \widehat{\Theta}_{a,s,s'} \left[\sum_{s^*} \left|\widehat{\Theta}^{a,s\to s'}_{a,s,s^*} - \widehat{\Theta}_{a,s,s^*}\right|\right]$$