universität innsbruck

An information-theoretical approach to internal models in a Partially Observable Markov Decision Process

By Lukas Prader

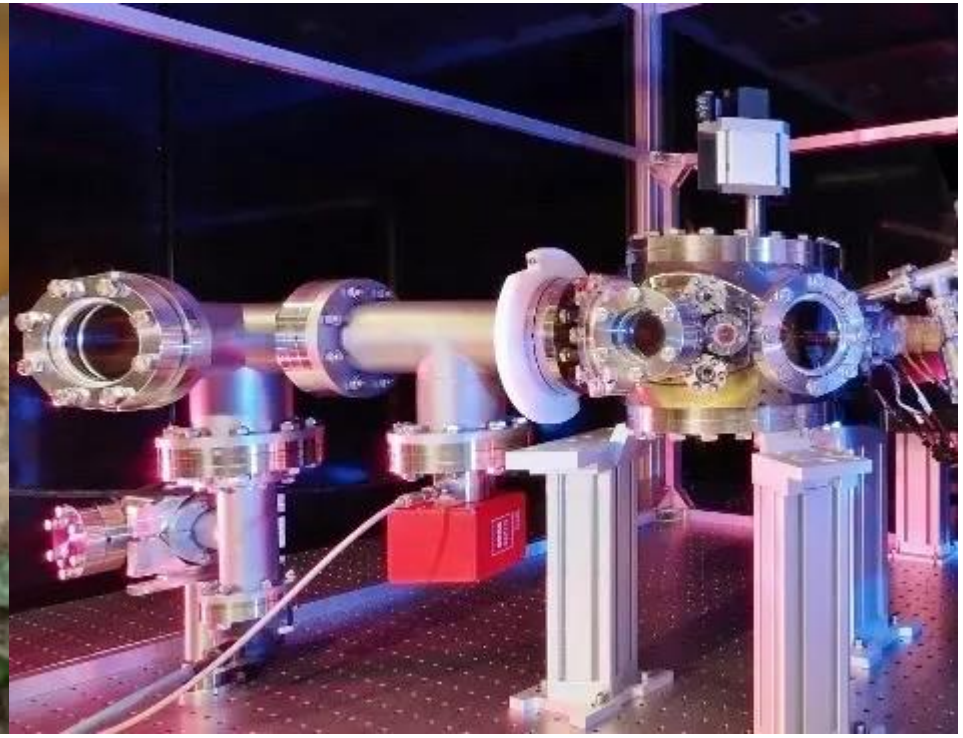
Supervisors: Alexander Vining, Hans J. Briegel

Group: Quantum Information & Computation

How can an agent learn about the structure of its environment?

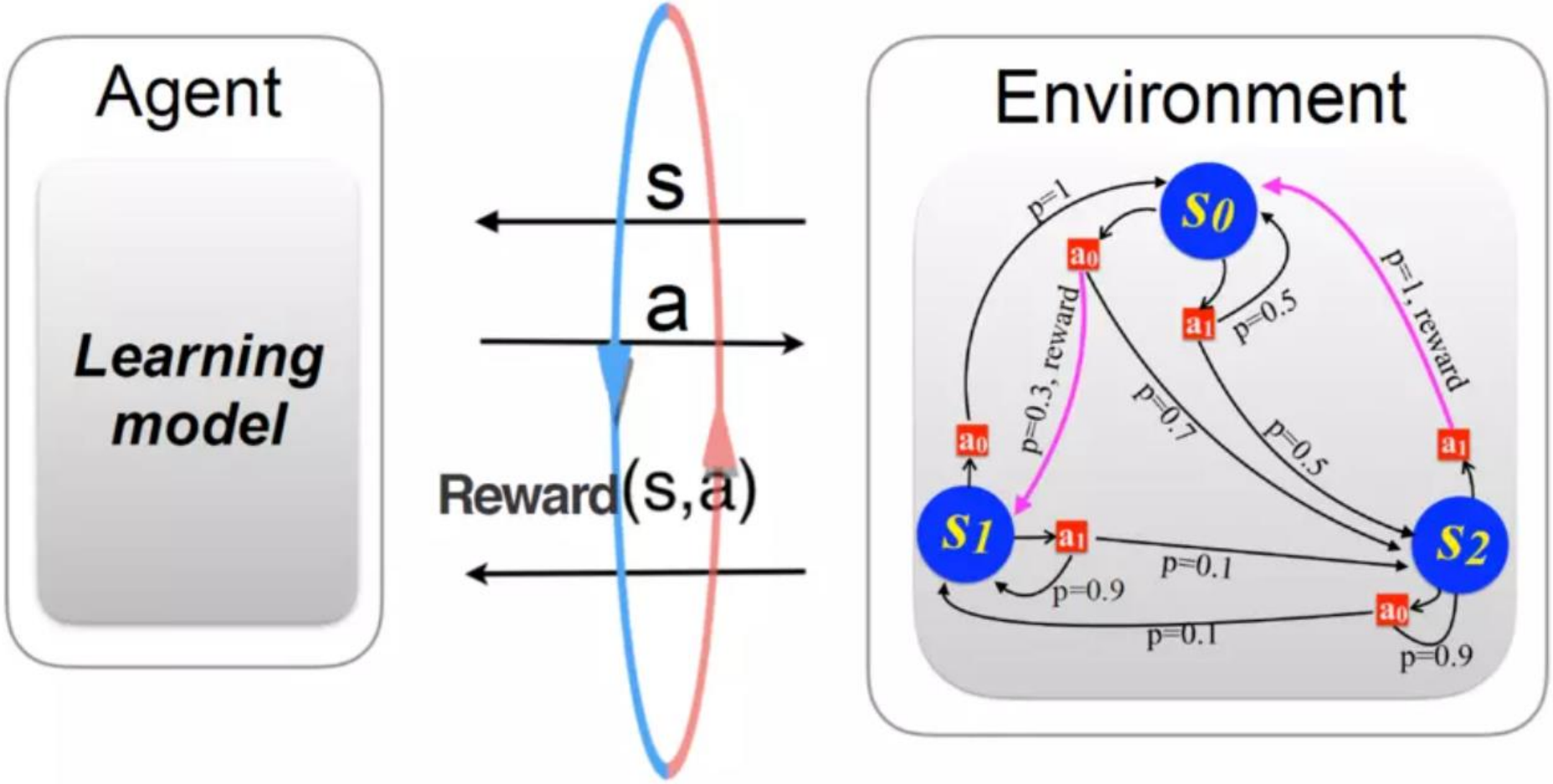
https://www.westend61.de/en/photo/FO006476/brazil-mato-grosso-mato-grosso-do-sul-pantanal-giant-anteater-and-termite-hill

https://www.livescience.com/64104-how-do-squirrels-find-buried-nuts.html



https://www.uibk.ac.at/exphys/index.html.en

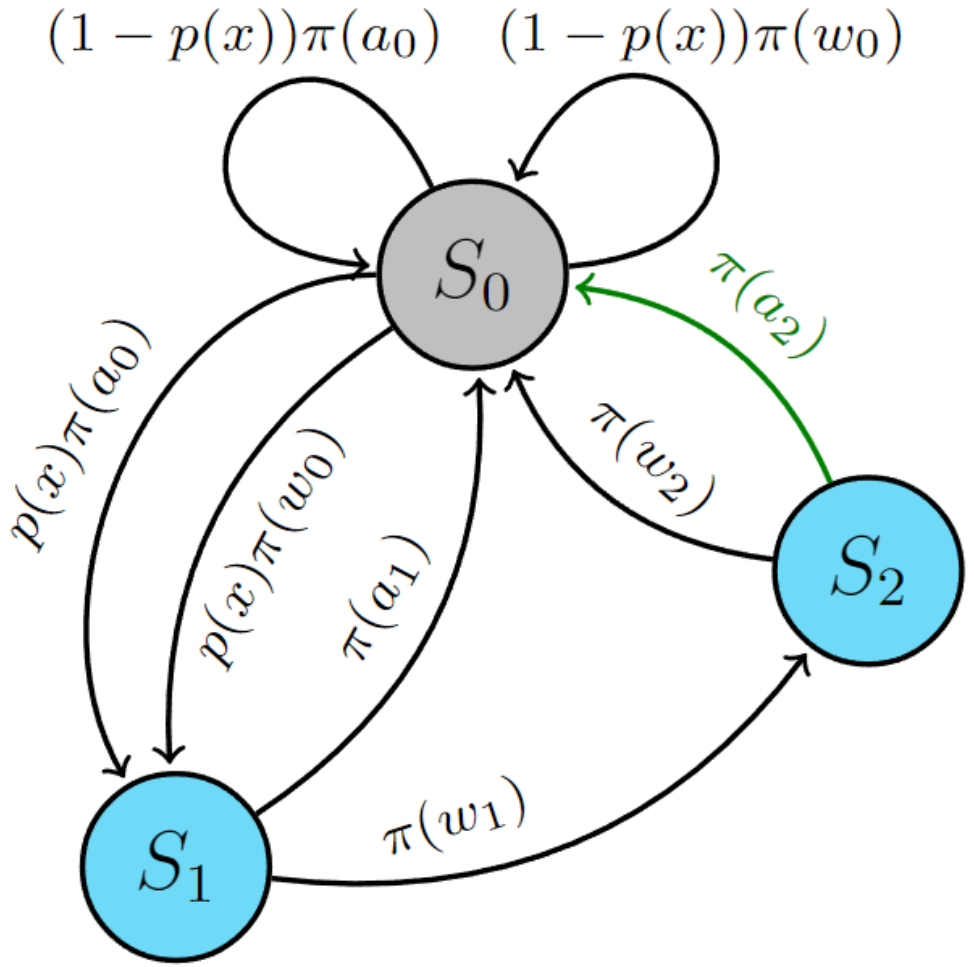
How can an agent learn about the structure of its environment?



https://www.uibk.ac.at/th-physik/qic-group/research/topics/ai-and-science/

The Skinner Box as a Partially Observable Markov Decision Process

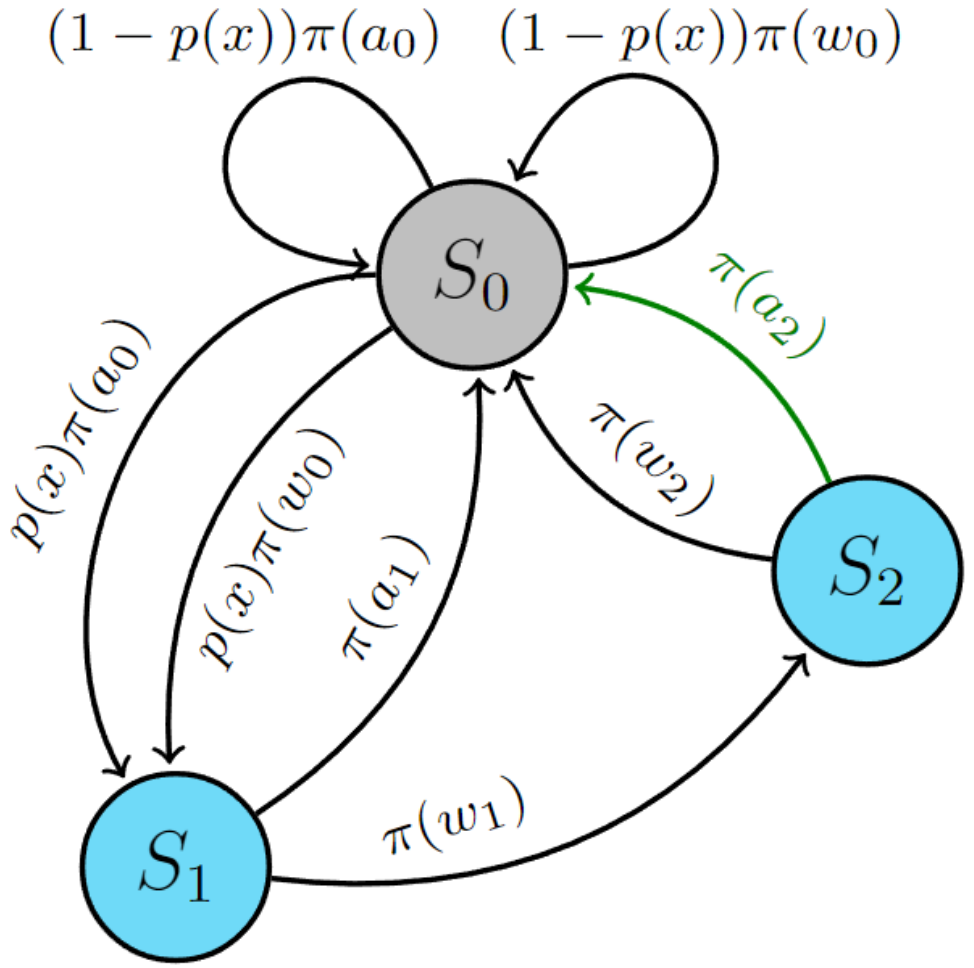
https://www.youtube.com/watch?v=lxImLwqY0Jc&t=1s



The Skinner Box as a Partially Observable Markov Decision Process cont.

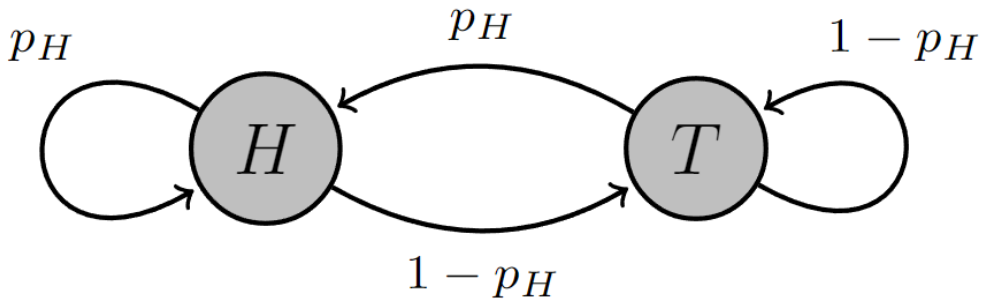
What observations are needed to understand this process?

While only observing the light, can an observer infer the existence of “hidden” states?



Markov Processes and Information Theory

- Markov property
- **State transition matrix** P
- **Stationary distribution** μ:



Markov Process for a biased coin

Use in physics:

- Brownian motion
- Micro-state transitions (statistical physics)

- **Shannon Entropy** $H = -\sum_i p_i \log p_i$

Cover & Thomas, Elements of Information Theory

Information Theory and Synchronization

- **Block entropy** $H(L)$

i.e. coin flip: $\mathrm{L} = 2 \rightarrow$ (HH), (HT), (TH), (TT)

- **Entropy rate** h_μ

How does the block entropy increase with L in the limit?

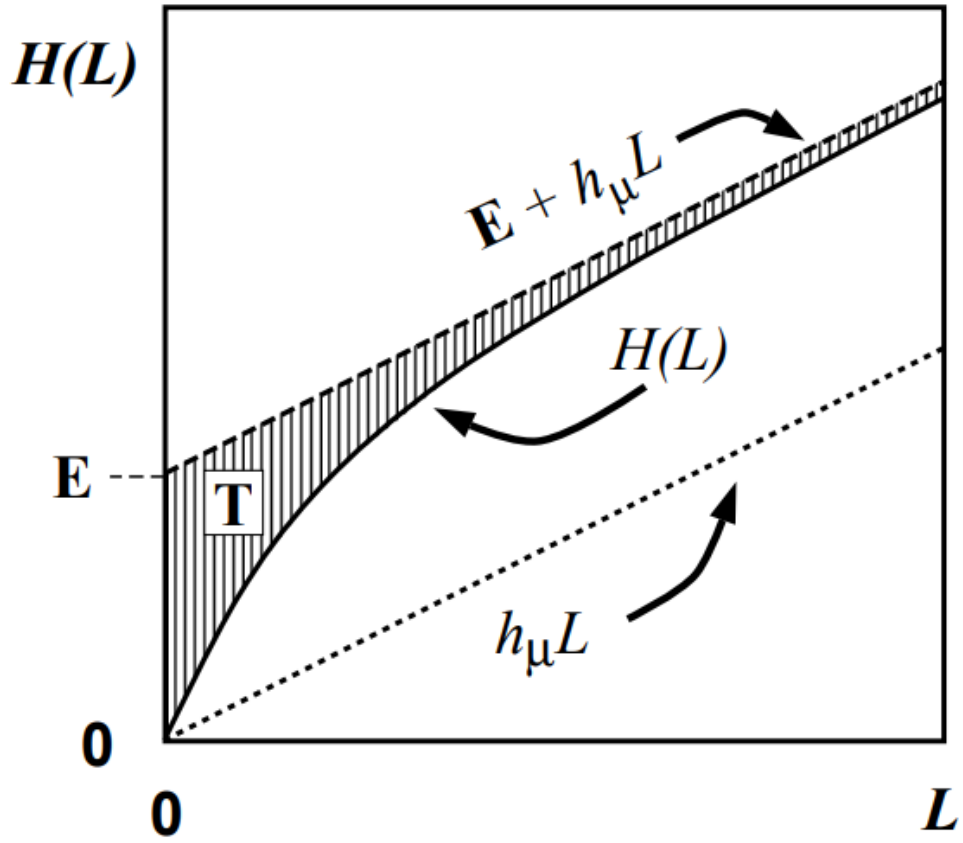
- **Excess entropy** E

How much information has to be learned to synchronize?

- **Transient information** T

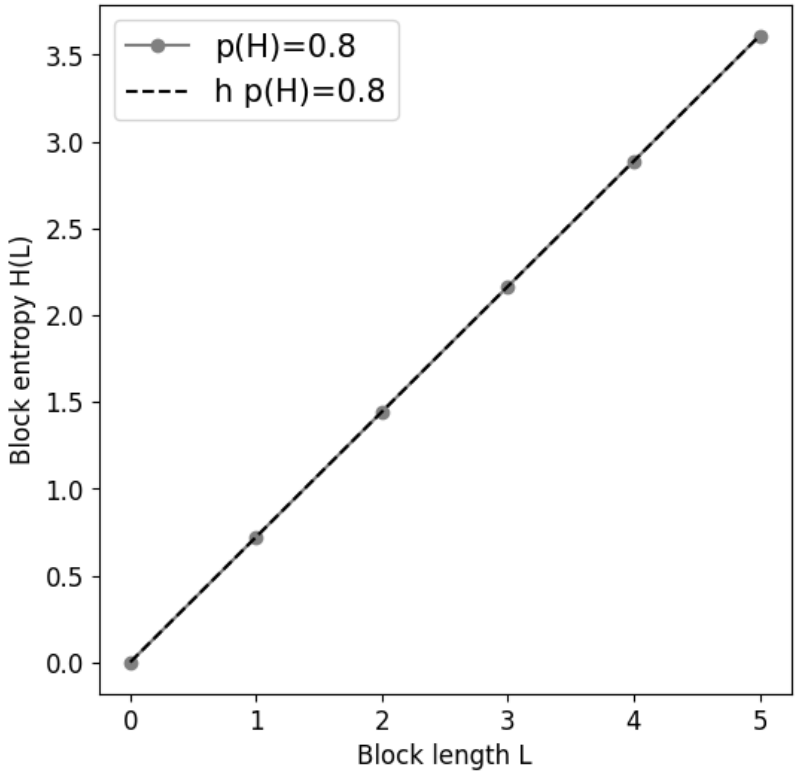
How much uncertainty does the agent encounter during synchronization?

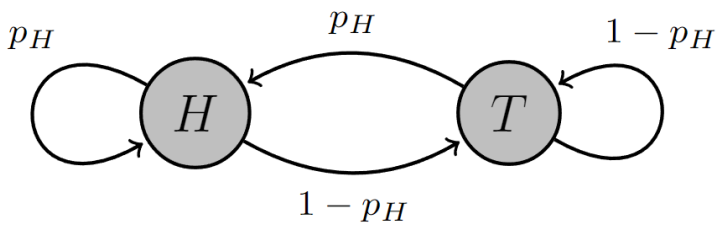
Learning as “synchronization” to the environment

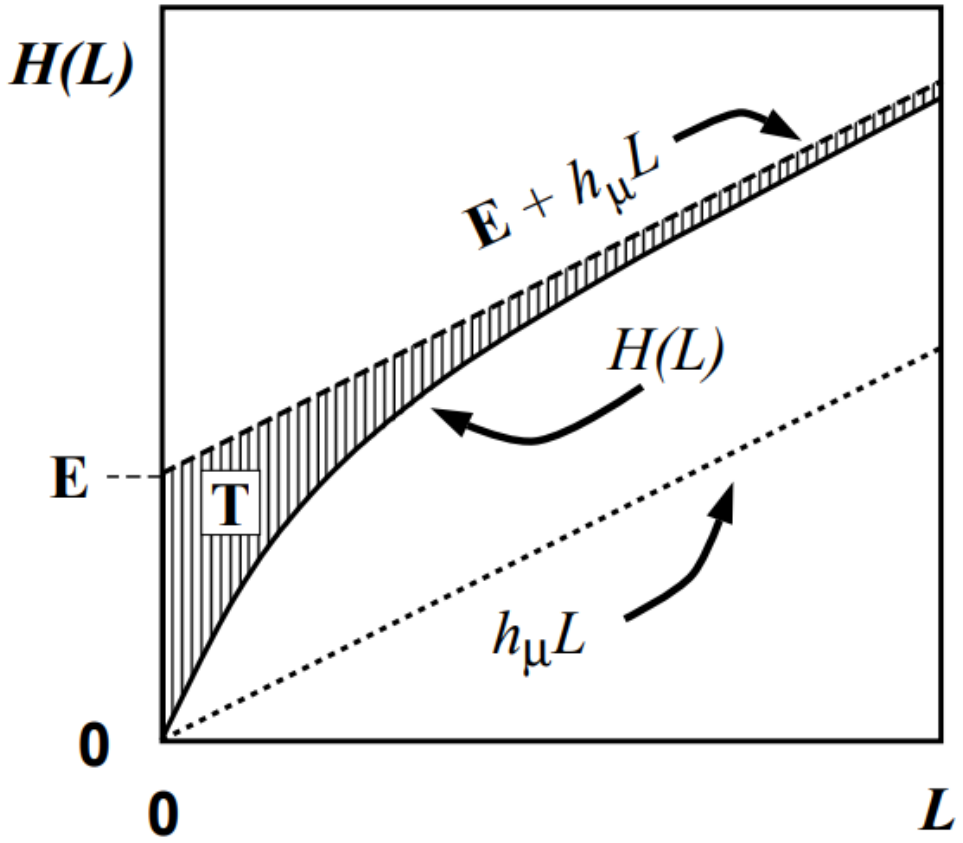


Crutchfield & Feldman, 2003

Information Theory and Synchronization for a biased coin



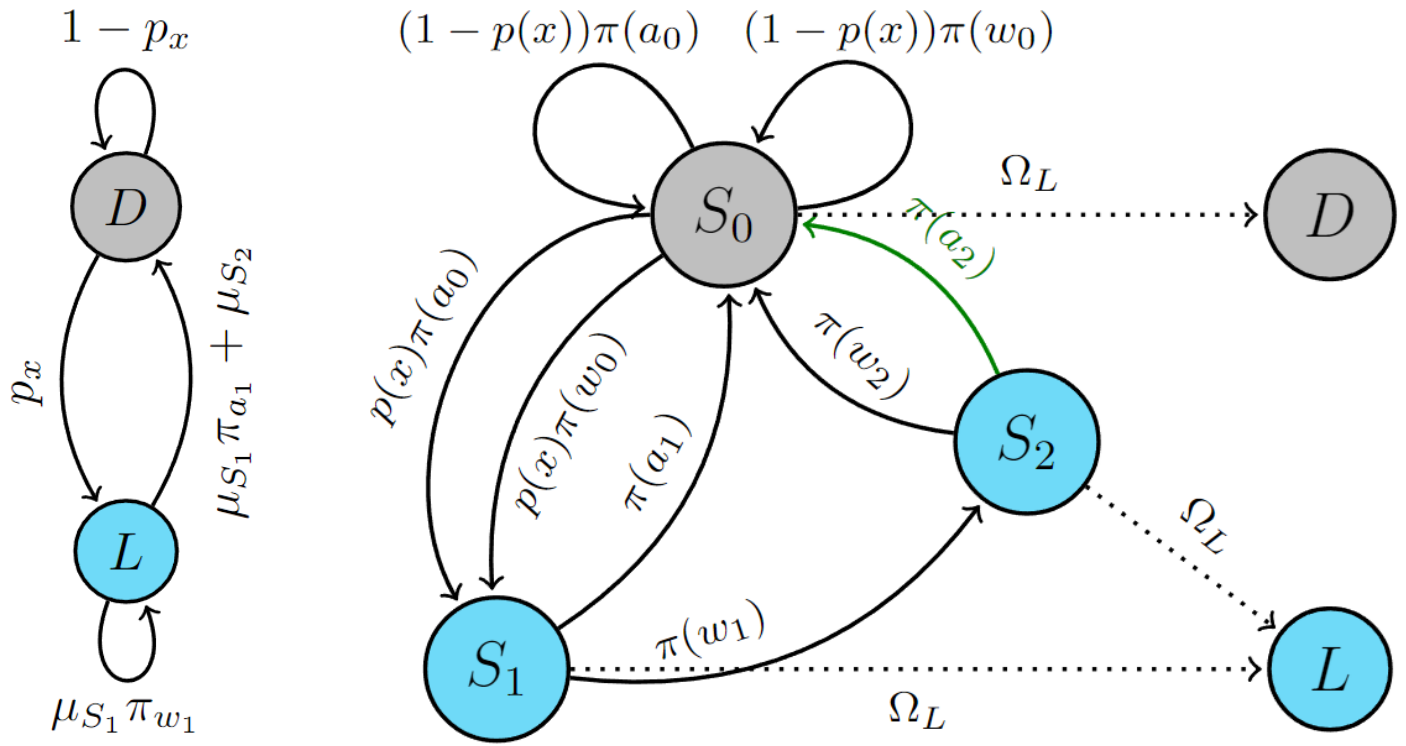




Crutchfield & Feldman, 2003

Reducing the delayed action MDP

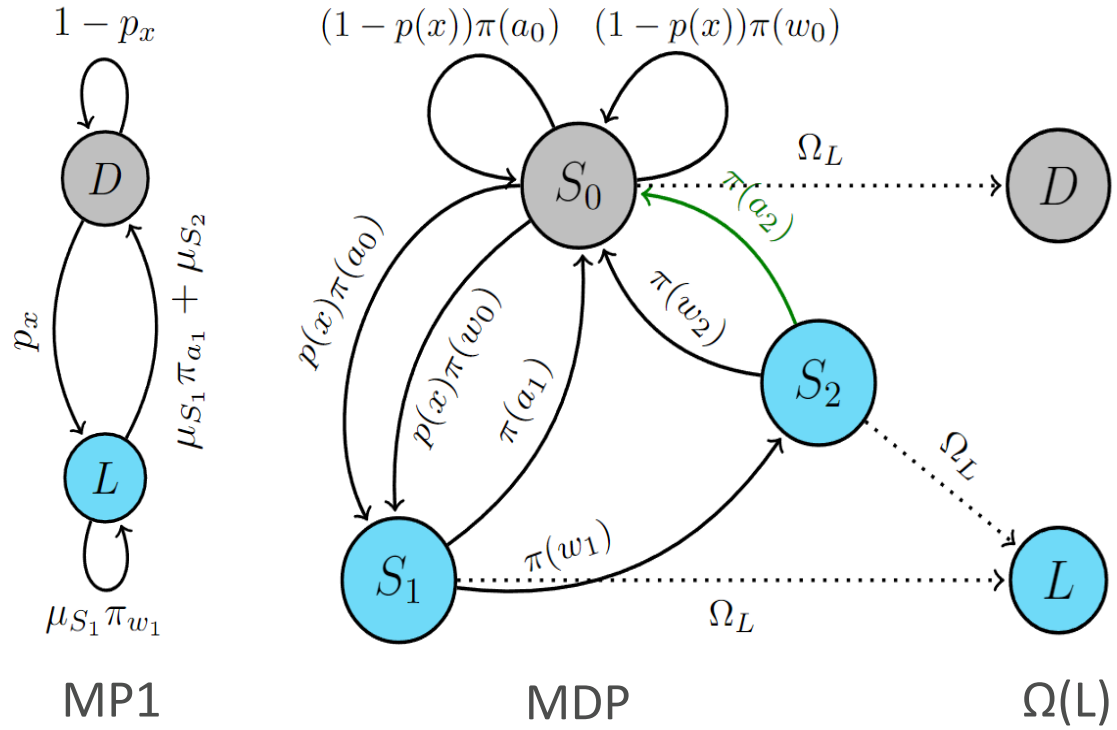
- Only parts of the process can be observed
- The process is assumed to already be in stationarity

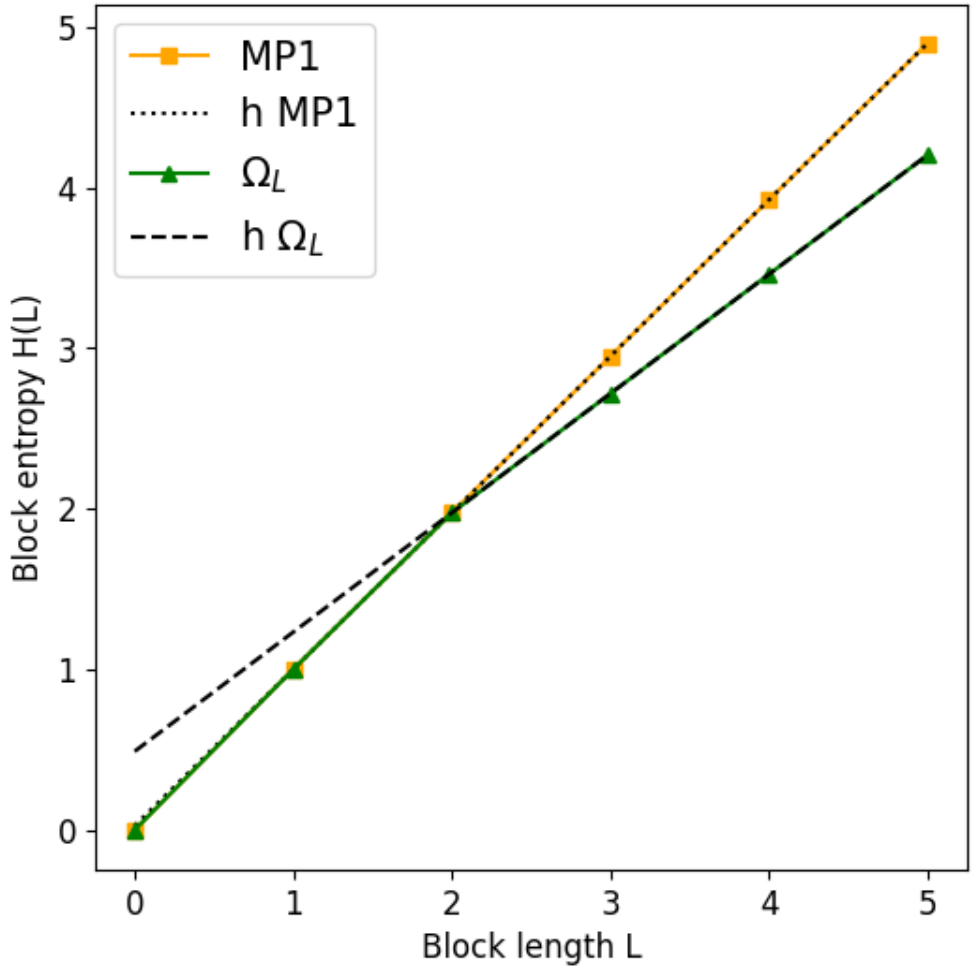


$S_0S_0S_1 \rightarrow DDL$
$S_1S_0S_1 \rightarrow LDL$
$S_2S_0S_1 \rightarrow LDL$
$\vdots$

MP1 MDP Ω(L)

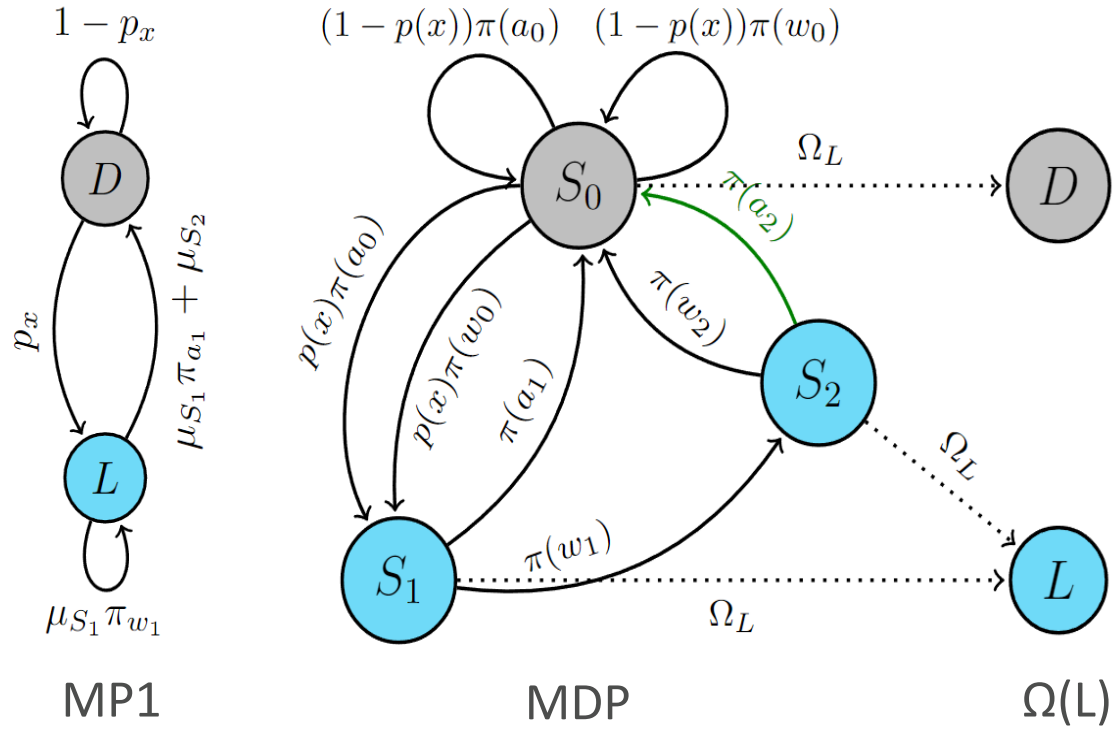
Analysing the block entropy behaviour

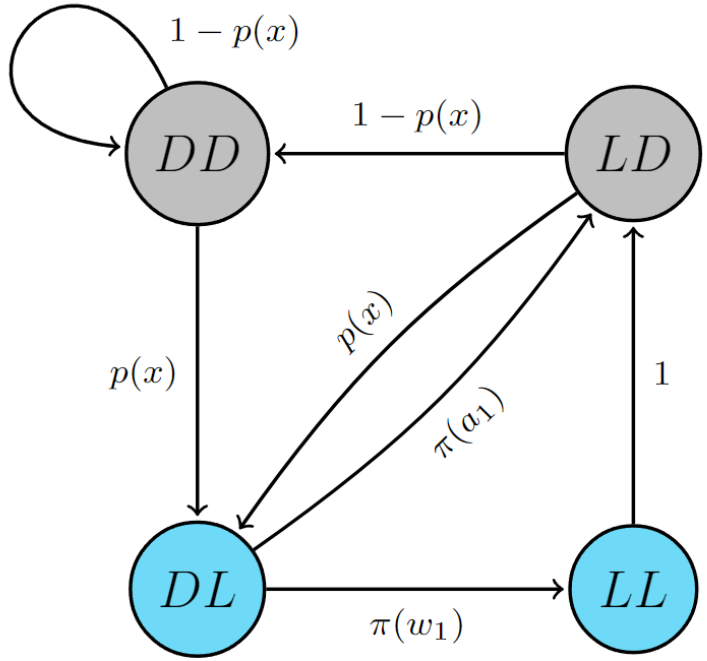




- Parameters: $[\mathrm{p(x)}, \pi(a_0), \pi(a_1), \pi(a_2)]$
  $[\ 0.6\ ,\ 0.5\ ,\ 0.3\ ,\ 0.7\ ]$
- $E \approx 0.49$ bits $T \approx 1.72$ bits·symbols

Analysing the block entropy behaviour



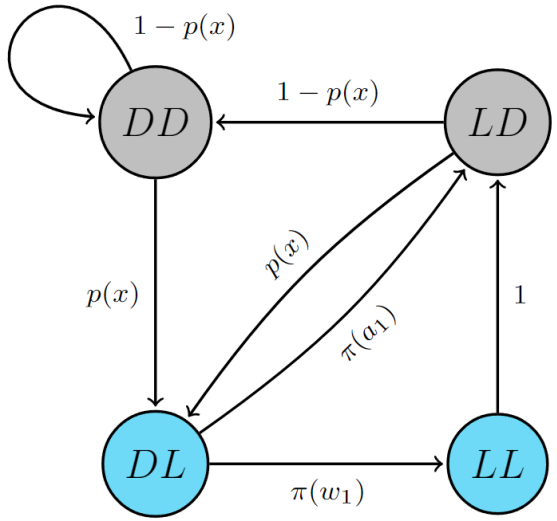


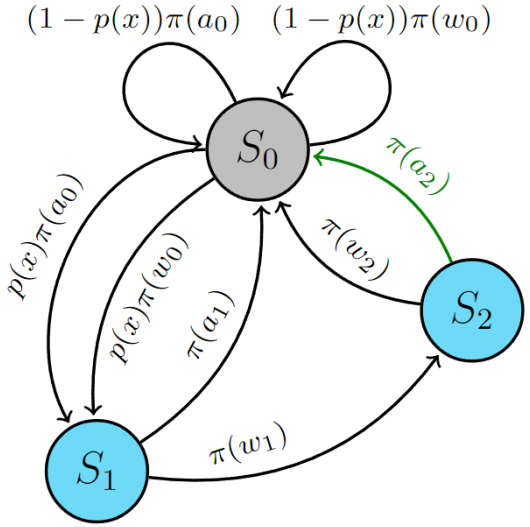
Future directions

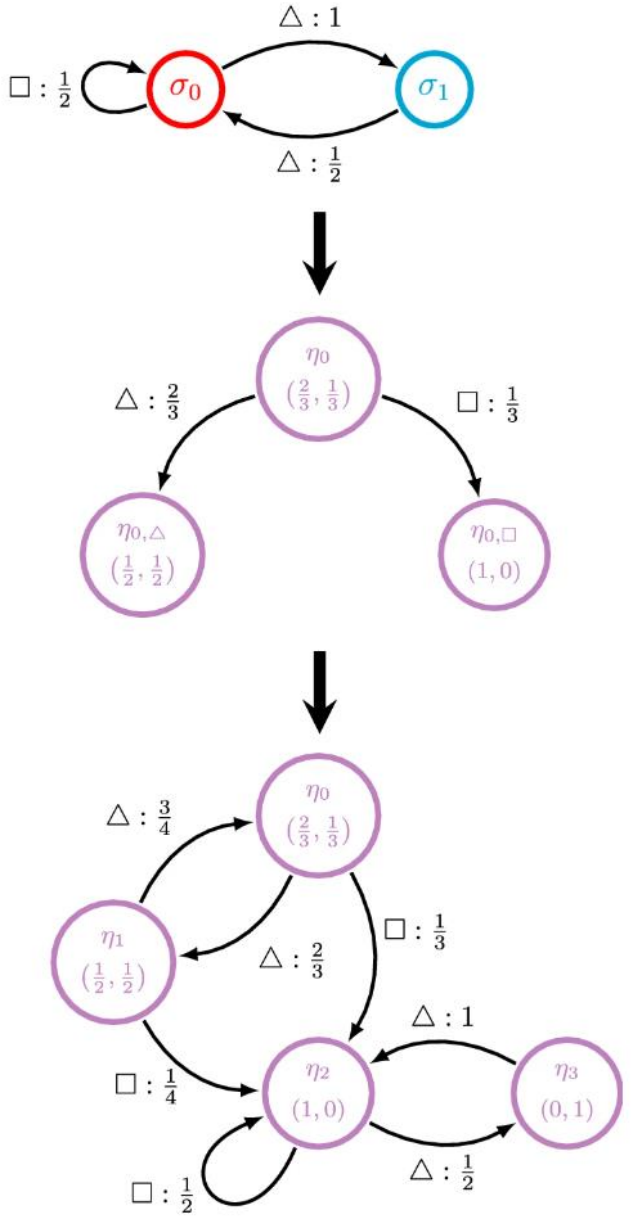
- Look at observation of the action instead of light
- Uncertainty bounds and minimum sample length
- Use for internal model generation
- How can an agent associate actions with observations in this model?

Use for internal model generation

- Synchronization as an indicator for sufficient information
- Creation of an “ϵ-machine” using the recurrent part of a mixed-state system
- Using the information to optimize the policy, relating to reward







Jurgens & Crutchfield, 2021

Conclusion

- The necessary length of observations for understanding the light behaviour of the delayed action task is L=2.

- Information theoretical measures can be used to understand the process of internal model formation.

Literary sources

- Cover & Thomas: *Elements of Information Theory*, 1999
- Crutchfield & Feldman: *Regularities Unseen, Randomness Observed: Levels of Entropy Convergence*, 2003
- Jurgens & Crutchfield: *Shannon entropy rate of hidden Markov processes*, 2021

/uniinnsbruck

www.uibk.ac.at