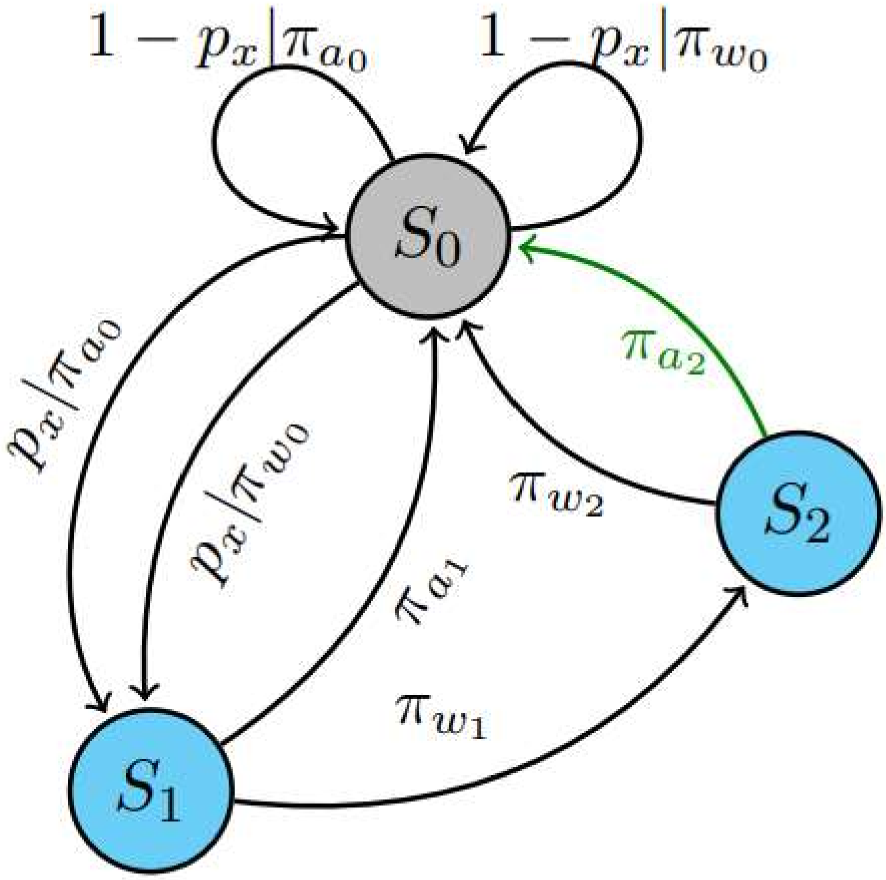
An information-theoretical approach to internal models in a Partially Observable Markov Decision Process

Rat in a box



MDP
$p_x = \text{probability of light turning on}, \; \pi_{a_i}, \pi_{a_i} = \text{agent policy to act or wait}$

The rat can act (push the button) or wait, with reward for acting in S_2

How can an observer see that there is a more complex process going on?

Measures from information theory, (Crutchfield & Feldman, 2001):

Shannon entropy for blocks of some length L, where each block is a sequence of L consecutive observations of $s_i \epsilon S$:

$$H(s^L) = -\sum_{s_i^L \epsilon S^L} p(s_i^L) \log_2 p(s_i^L)$$

Example for MDP, $L = 1, 2$:

$$H(s^1) = -[p(S_0)\log_2(S_0) + p(S_1)\log_2(S_1) + p(S_2)\log_2(S_2)]$$

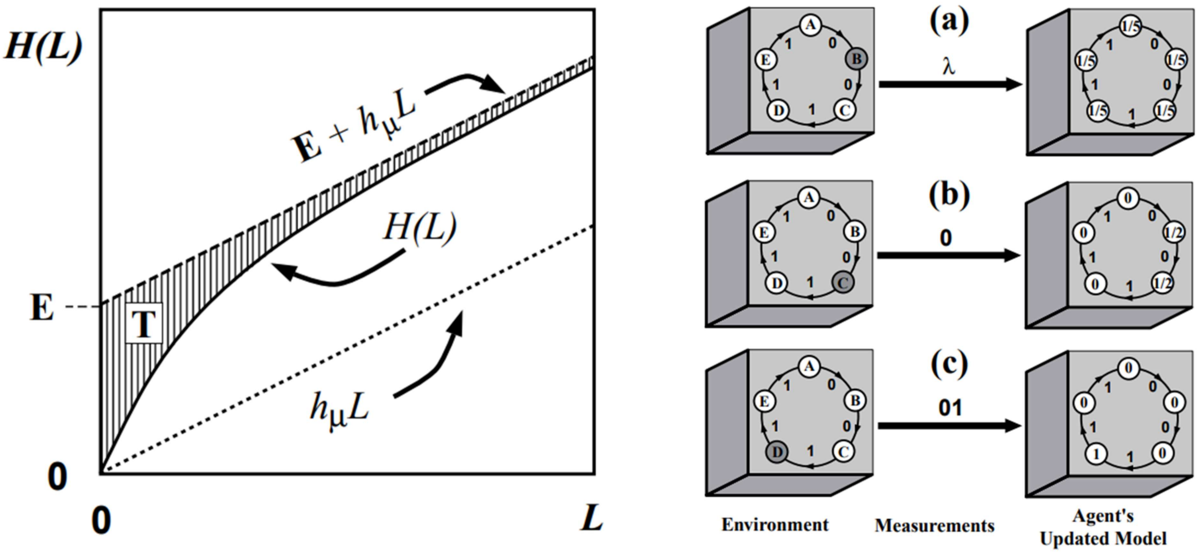
$$H(s^2) = -[p(S_0S_0)\log_2(S_0S_0) + p(S_0S_1)\log_2(S_0S_1) + p(S_1S_0)\log_2(S_1S_0) + p(S_1S_2)\log_2(S_1S_2)]$$

How does this block entropy change for longer and longer sequences? $\rightarrow$ **entropy rate** h_μ

$$h_\mu = \lim_{L \to \infty} \frac{H(s^L)}{L}$$

How does the block entropy $H(s^L)$ **converge to the entropy rate** h_μ**?** $\rightarrow$ **excess entropy** E

$$E = \lim_{L \to \infty} [H(s^L) - h_\mu L]$$



Figures from (Crutchfield & Feldman, 2001). Left showing the block entropy $H(s^L)$, here $H(L)$, behaviour for increasing L, Right showing concept of synchronization.

When is an agent able to estimate the entropy rate h_μ **correctly?** $\rightarrow$ **synchronization**

The agent is said to be synchronized, if block entropy scales with entropy rate, so
$E + h_\mu L - H(s^L) = 0$
The total uncertainty encountered until this synchronization is achieved is called transient information T

$$T = \sum_{L=0}^{\infty} E + h_\mu L - H(s^L)$$

So, what about the rat problem?

(Crutchfield & Feldman, 2001) mainly looks at periodic sequences, what about a Markov Process?

Additional definitions for MPs (Cover & Thomas, 1999):

Given a MP with state-transition matrix P, if the MP is stationary (transition probabilities do not change), then: μ is the stationary distribution vector with:

$$\mu = \mu P$$

$\rightarrow$ solve as a special eigenvector problem.

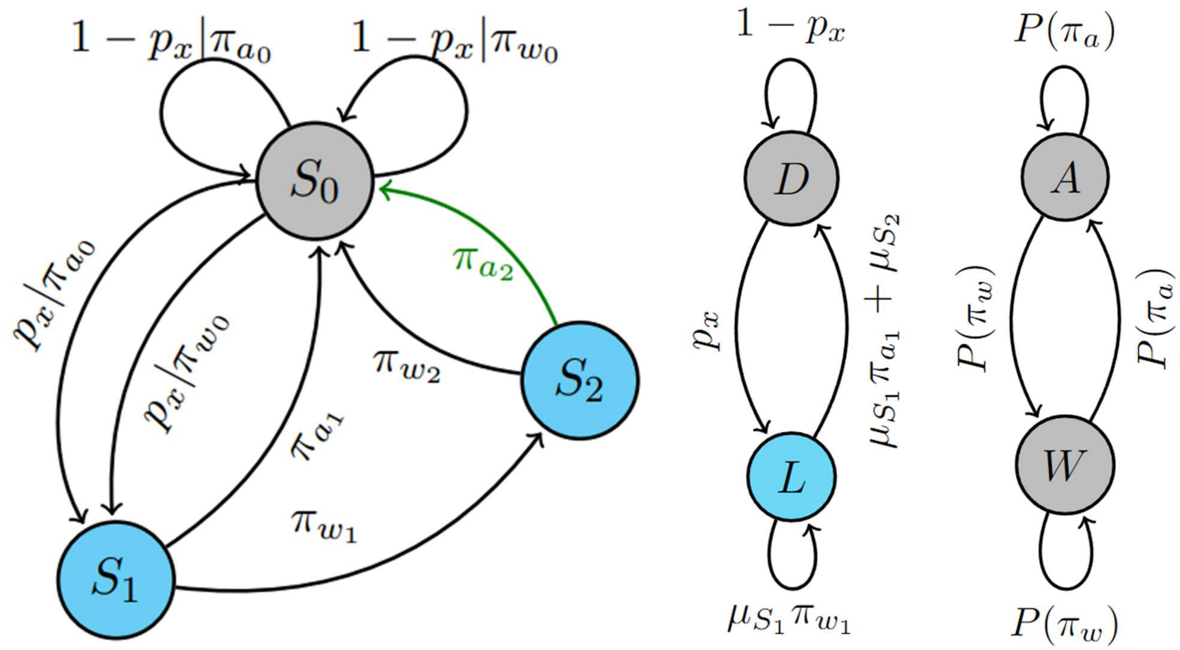
$\implies$ the MP has an entropy rate given by:

$$h_\mu = \sum_{i,j} \mu_i P_{ij} \log_2 P_{ij}$$

For the rat in the box:

How can we model this synchronization for an observer of the system?
Assumption: we start to observe the system in stationarity.

First approach, create observation Markov processes



MDP

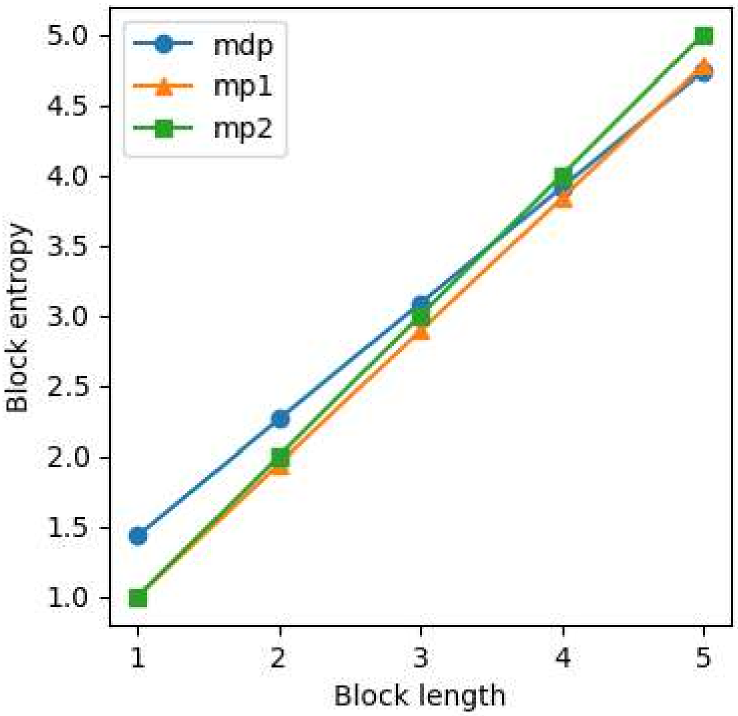
$$\begin{bmatrix} \frac{1}{-\pi_{a1} p_X + 2p_X + 1} & \frac{p_X}{-\pi_{a1} p_X + 2p_X + 1} & \frac{p_X(\pi_{a1} - 1)}{\pi_{a1} p_X - 2p_X - 1} \end{bmatrix}$$

MP1

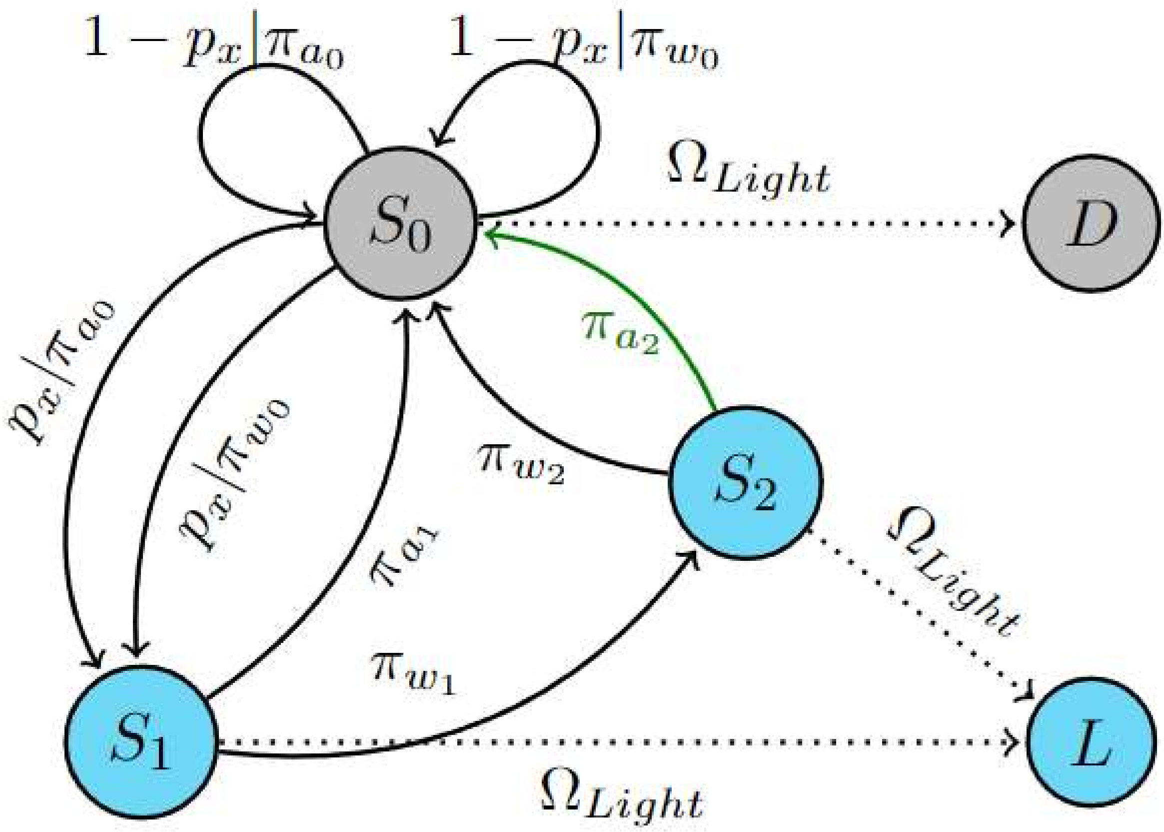
$$\begin{bmatrix} \frac{1}{-\pi_{a1} p_X + 2p_X + 1} & \frac{p_X(\pi_{a1} - 2)}{\pi_{a1} p_X - 2p_X - 1} \end{bmatrix}$$

MP2

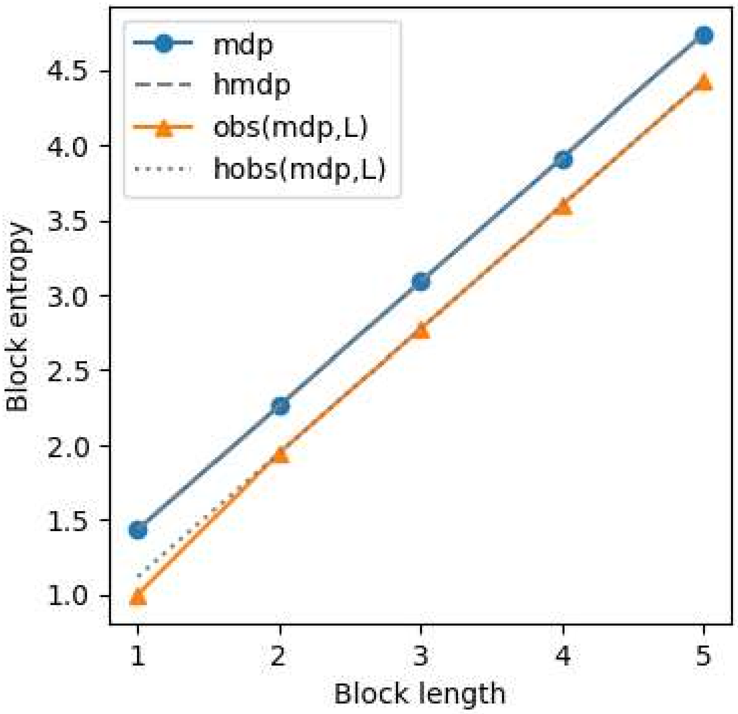
$$\begin{bmatrix} \frac{\pi_{a0} + \pi_{a1} p_X - \pi_{a2} p_X (\pi_{a1} - 1)}{-\pi_{a1} p_X + 2p_X + 1} & \frac{-\pi_{a0} - 2\pi_{a1} p_X + \pi_{a2} p_X (\pi_{a1} - 1) + 2p_X + 1}{-\pi_{a1} p_X + 2p_X + 1} \end{bmatrix}$$



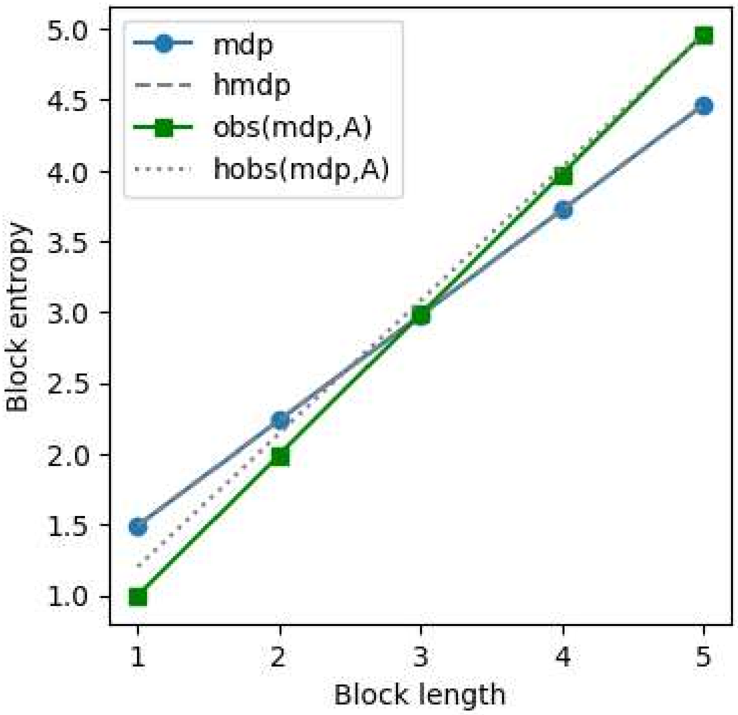
Second approach, treat it as a Hidden Markov Process



```
mdp
entropy rate
analytical | discrete estimate
0.826816102344562 | 0.573105250181697/log(2)
[0.993268210149961/log(2), 1.56637346033166/log(2), 2.13947871051335/log(2), 2.71258396069
505/log(2), 3.28568921087675/log(2)]
obs(mdp,L)
entropy rate
analytical | discrete estimate
0.826816102344562 | 0.573105250181697/log(2)
[0.691761498852418/log(2), 1.34748487699694/log(2), 1.92059012717863/log(2), 2.49369537736
033/log(2), 3.06680062754203/log(2)]
```




```
mdp
entropy rate
analytical | discrete estimate
0.742438185145091 | 0.514618934773364/log(2)
[1.03519929987627/log(2), 1.54981823464963/log(2), 2.064437169423/log(2), 2.5790561041963
6/log(2), 3.09367503896972/log(2)]
obs(mdp,A)
entropy rate
analytical | discrete estimate
0.940057780799657 | 0.685278465029803/log(2)
[0.692511814145032/log(2), 1.38144476798026/log(2), 2.06907890375316/log(2), 2.75464303978
077/log(2), 3.43992150481057/log(2)]
```


Block entropy
Block length
mdp
hmdp
obs(mdp,A)
hobs(mdp,A)
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
1
2
3
4
5