

Interactive learning

Susanne Still

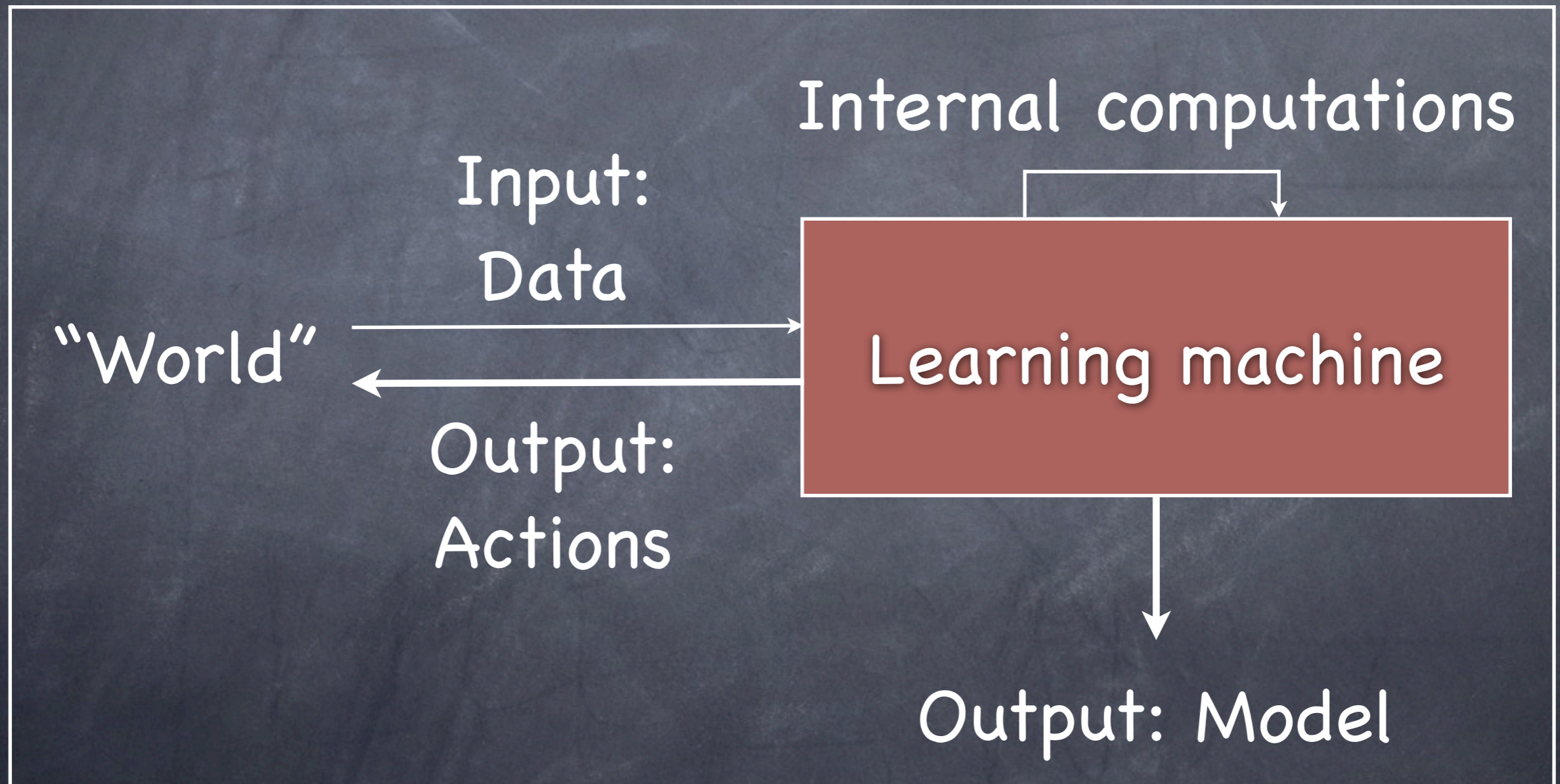
University of Hawaii at Manoa

Honolulu, HI

sstill@hawaii.edu

Information theoretic approach to interactive learning. EPL 2009
on my website and at <http://arxiv.org/abs/0709.1948>

The challenge



Motivations

- Come up with theoretical guidance for how to measure/perturb a physical system, such that
- The measurements/perturbations aid model making.
- Side product: Theoretical description of automated model making (online-OCI).
- Possible applications:
 - data rich domains, such as astronomy, particle physics, climate models, molecular biology, genomics
 - quantum information theory, quantum measurement
 - Robotics and AI

Problem

An agent observes an environment, and uses those observations to build a model.

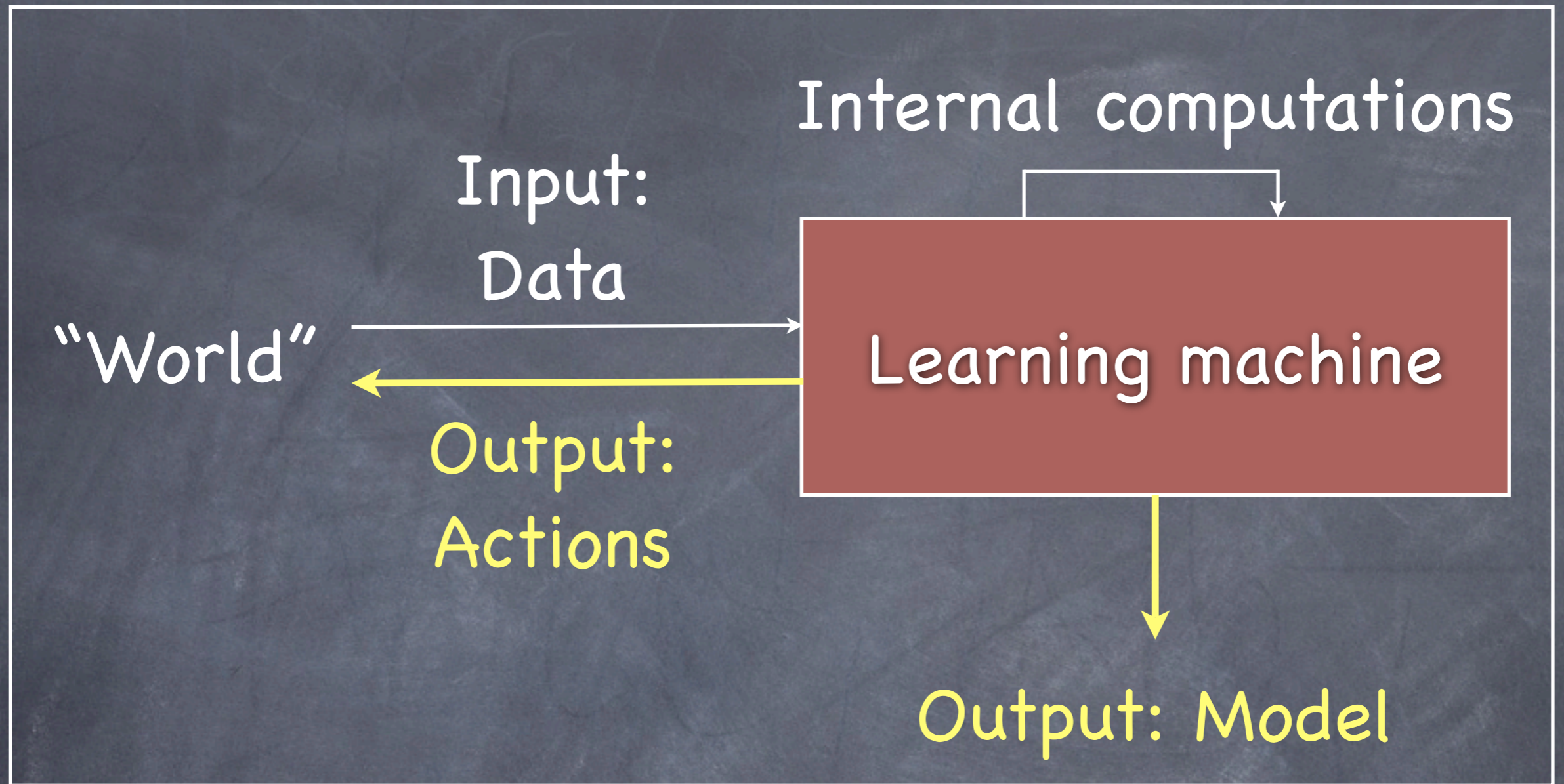
The agents also acts on the environment, and its actions influence the data it observes.

What is a good action strategy?

How to act on the world?

What is an appropriate model, in the presence of the feedback?

What to think about the world?



● Fundamental questions:

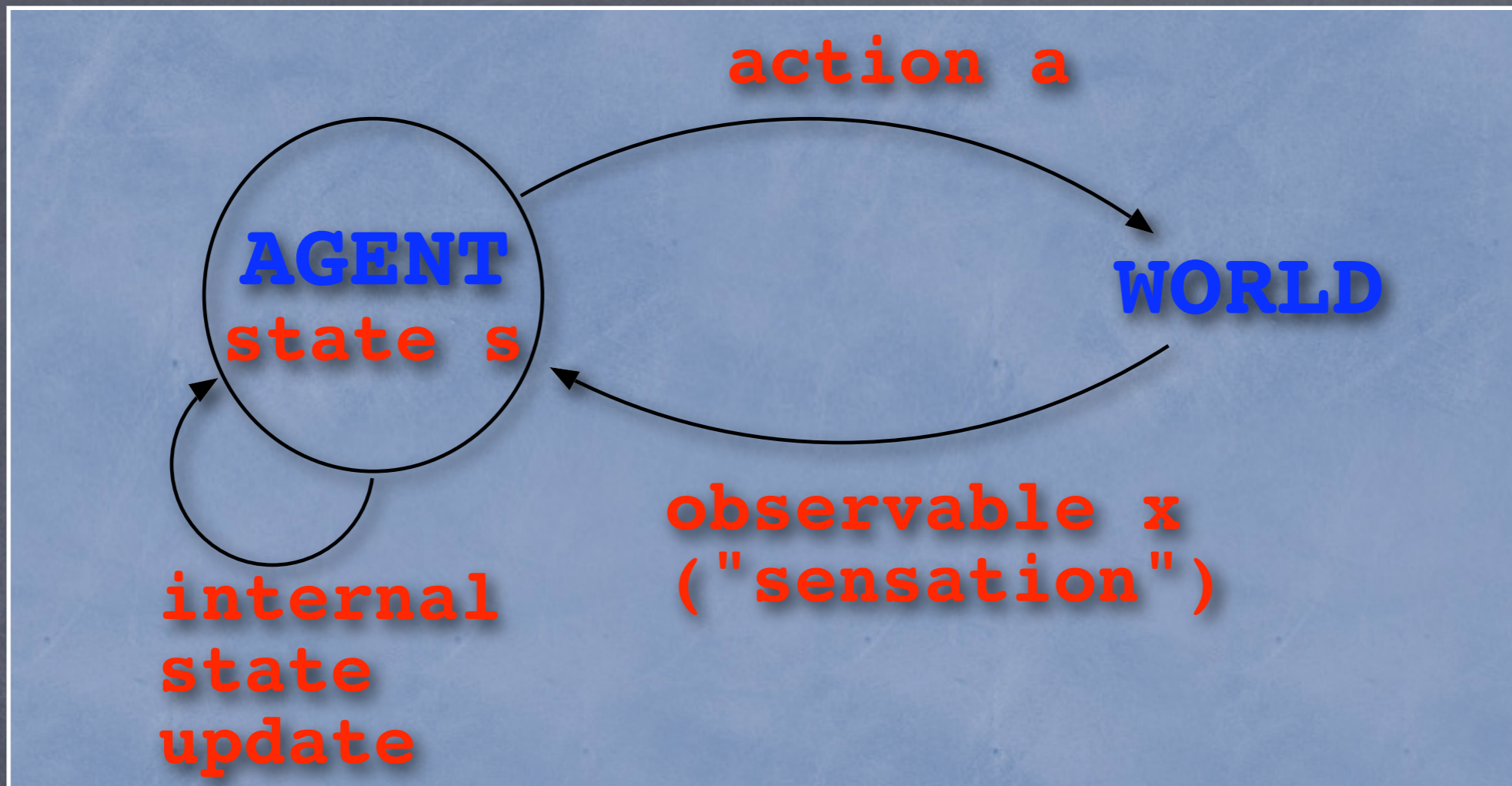
- What is a "good" model?
- What is an appropriate action policy for learning such a model?

What is a good model?

1. A good model predicts well
 - > find maximally predictive model.
Keep predictive information!
2. A good model is compact
 - > do not keep irrelevant information.

What is an appropriate action policy?

- Action policy influences future data
- Predictive information of the observed time series depends on the action policy
=> Maximally achievable predictive power of the model depends on the action policy
- Find the action strategy that allows for making a maximally predictive model.



Observed time series:

$$\dots, \overbrace{x_{t+T_x}, \dots, x_t}^{x_t^{\text{past}}}, \dots$$

Sequence of actions:

$$\dots, \overbrace{a_{t+T_a}, \dots, a_t}^{a_t^{\text{past}}}, \dots$$

Agent's internal states:
(internal representation)

$$\dots, s_{t+T_x}, \dots, s_t, \dots$$

$$\overbrace{x_{t+1}, \dots, x_{t+T'_x}}^{Z_t}, \dots$$

a_{t+1}, \dots
action changes future

s_{t+1}, \dots
state predicts future

Notation: Write all available information into one vector $v_t = (x_t^{\text{past}}, a_t^{\text{past}}, s_t)$

• The model **summarizes** the available information.

• The action is chosen in response to the available information.

• Want to find the maps: $v_t \mapsto s_{t+1}$

$v_t \mapsto a_{t+1}$

• These do not have to be deterministic, in general they can be probabilistic. Find:

$$p(s_{t+1}|v_t)$$

Model

$$p(a_{t+1}|v_t)$$

Action Policy

Quantifying the intuition

Optimization principle:

$$\max_{\substack{p(s_{t+1}|v_t) \\ p(a_{t+1}|v_t)}} (I[\{s_{t+1}, a_{t+1}\}; z_t] - \lambda I[s_{t+1}; v_t] - \mu I[a_{t+1}; v_t])$$

max predict. info in
presence of action

min model
complexity

find simple
action policy

Optimal Model Class

$$p(s_{t+1}|v_t) = \frac{P(s_{t+1})}{Z_S(v_t, \lambda)} e^{-\frac{1}{\lambda} E_S(s_{t+1}, v_t)}$$

$$E_S(s_{t+1}, v_t) = \left\langle D_{KL} \left[\frac{P(z_t|v_t, a_{t+1})}{P(z_t|s_{t+1}, a_{t+1})} \right] \right\rangle_{P(a_{t+1}|v_t)}$$

- Most likely state minimizes relative entropy between actual and predicted conditional future distribution, averaged over action policy
=> states capture effect that the past has on the future, under in the presence of actions.
- optimal model reflects the causal structure of the process.

Recall

- In the absence of actions, the learning procedure recovers in the limit $\lambda \rightarrow 0$; $t \rightarrow \infty$, the ϵ -machine:
- Hidden Markov model with the following properties:
 - Hidden states = causal states
 - Causal state partition: two histories are equivalent when $p(\text{future}|v_1) = p(\text{future}|v_2) =: p(\text{future}|s)$
 - Causal shielding: Conditional independence of future and past, given the causal states.
 - Deterministic Transitions $p(s'|x,s)$
 - Optimal predictor (all predictive information is kept)
 - Minimal size (smallest statistical complexity)
 - Unique. Sufficient statistics.

(J. P. Crutchfield and K. Young (1989) PRL 63:105–108)

Recall

- In general (finite λ) extension of ϵ -machine to non-deterministic models (can be more compact)
- Class of models that optimally trade complexity for prediction error.

S. Still, J. P. Crutchfield. Structure or Noise?
<http://lanl.arxiv.org/abs/0708.0654>

S. Still, J. P. Crutchfield, C. J. Ellison. Optimal Causal Inference.
<http://lanl.arxiv.org/abs/0708.1580>

Extension of causal states to interactive learning

- Optimal model for deterministic decisions ($\lambda \rightarrow 0$; $\mu \rightarrow 0$)

$$s_{t+1}^*(v_t) := \arg \min_s D_{KL} \left[\frac{P(z_t | v_t, a_{t+1}^*(v_t))}{P(z_t | s, a_{t+1}^*(v_t))} \right]$$

- Now we can define a new equivalence class, extending the causal state partition to the situation with feedback from the learner:
- Two histories v and v' are causally equivalent under the deterministic action policy, $a = f(v)$, if $P(z|v,a) = P(z|v',a)$.

Causal Equivalence

- Two histories v and v' are causally equivalent under the deterministic action policy, $a = f(v)$, if $P(z|v,a) = P(z|v',a)$.
- This is a partition of the space of histories which **groups all pasts that are the same w.r.t. predicting the future, in the presence of the perturbation $a(t)$.**
- Choice of the policy determines the observable time series which is produced by the system **coupled** to the observer via the actions.
- There could be two different action policies which result in observations with the same underlying causal state partition.
- The optimal deterministic policy a^* creates that coupled system which can be predicted most effectively by a causal model!

Optimal Action policy

$$\pi(a_{t+1}|v_t) = \frac{P(a_{t+1})}{Z_A(v_t, \mu)} e^{-\frac{1}{\mu} E_A(a_{t+1}, v_t)}$$

$$E_A(a_{t+1}, v_t) = \left\langle D_{KL} \left[\frac{P(z_t|v_t, a_{t+1})}{P(z_t|s_{t+1}, a_{t+1})} \right] \right\rangle_{P(s_{t+1}|v_t)} - D_{KL} \left[\frac{P(z_t|v_t, a_{t+1})}{P(z_t)} \right]$$

Modeling accuracy

Information gain

Balance between control and perturbation!

Deterministic decisions:

$$\lambda \rightarrow 0; \mu \rightarrow 0$$

$$a_{t+1}^*(v_t) := \arg \min_a \left(D_{KL} \left[\frac{P(z_t | v_t, a)}{P(z_t | s_{t+1}^*(v_t), a)} \right] - D_{KL} \left[\frac{P(z_t | v_t, a)}{P(z_t)} \right] \right)$$

- Trade-off between control and exploration survives.
- Exploration is an optimal behavior. It is not equal to policy randomization!
- Compare to "Boltzmann Exploration" (reinforcement learning)

Example: A binary world of magnetic spins.

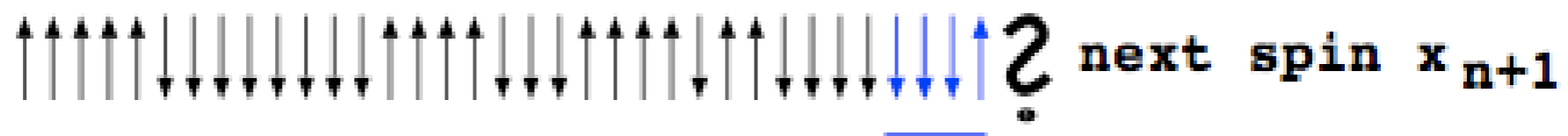
$$Q(x_{n+1}|a_{n+1}, \mathbf{w}_n) = \frac{e^{x_{n+1}(H(\mathbf{w}_n) + a_{n+1})}}{2 \cosh [H(\mathbf{w}_n) + a_{n+1}]}$$

Spin interaction:

$$H(\mathbf{w}_n) = \sum_{i=n-T+1}^n J_{n-i} x_i$$

observations: magnetic spins

\mathbf{w}_n (length T)



current record v_n (length L)

current internal state s_n



new action: external field



new internal state s_{n+1}

- Case L=T: Optimal state

$$s^*(\mathbf{v}_n) = \arg \min_{s_{n+1}} |s_{n+1} - H(\mathbf{v}_n)|$$

- Optimal action

- action is a small perturbation → reinforce belief!

$$a^*(\mathbf{v}_n) = \arg \max_{a_{n+1}} |a_{n+1} - H(\mathbf{v}_n)|$$

- large perturbation: trade-off with not creating too much order!

$$a^*(\mathbf{v}_n) = \arg \max_{a_{n+1}} \left(|a_{n+1} - H(\mathbf{v}_n)| + \sum_{x_{n+1}} Q(x_{n+1} | a_{n+1}, \mathbf{v}_n) \log [\langle Q(x_{n+1} | a_{n+1}, \mathbf{v}_n) p(\mathbf{v}_n) \rangle_{p(\mathbf{v}_n)}] \right)$$

Conclusions

- If an agent acts with the objective to learn an optimally predictive and compact model, then the resulting **optimal policy has to balance perturbation and control.**
- This follows purely from information theoretic principles.
- The balance is also present in the optimal deterministic policy. => Exploration is more than just policy randomization!
- Theory for optimal predictive inference in the presence of feedback from the learner.
- Extension of the notion of causal states and causal compressibility to modeling in the presence of feedback.
- Next Lecture: Fresh look at reinforcement learning.