Infomax Control: *A model of the real-time organization of behavior*

Paul Ruvolo

pruvolo@cs.brynmawr.edu

In search of a theory of behavior

Aims for the Behavioral Sciences:

Understand behavior
Predict behavior
Design interventions

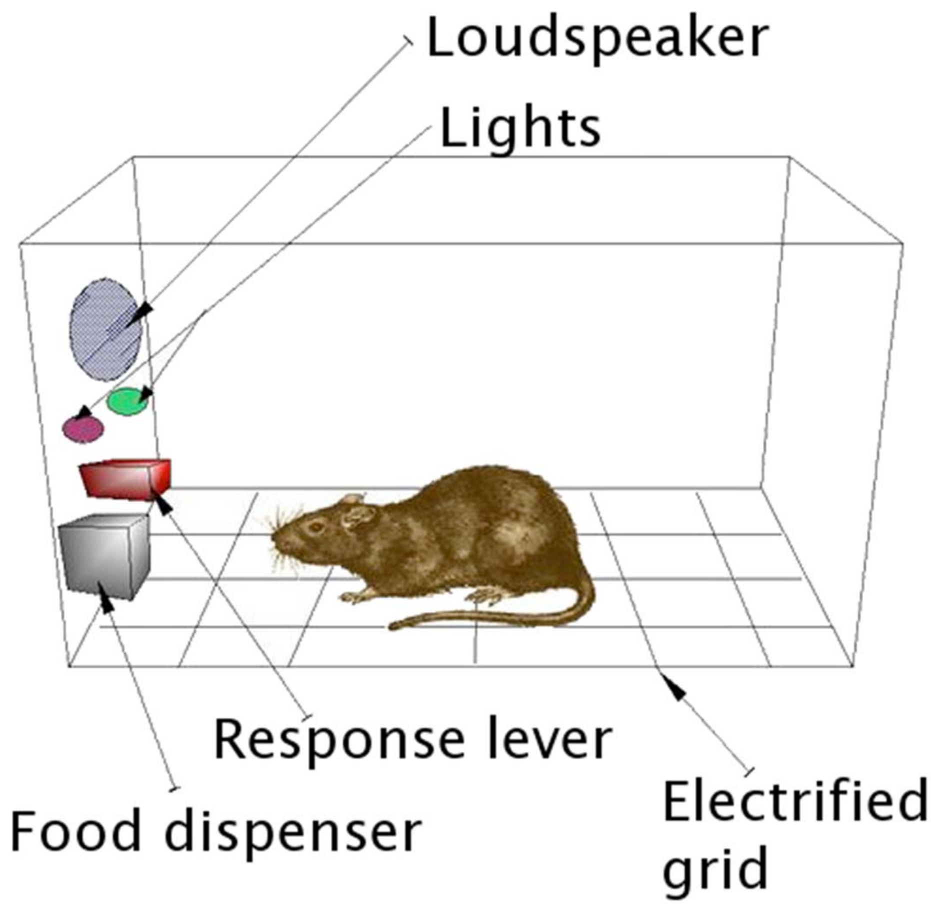
Actual baby, not a robot

Aims for Robotics:

Synthesize intelligent behavior for complex, unpredictable, and dynamic environments

Willow Garage's PR2 Robot

Skinnerian Behaviorism and Operant Conditioning



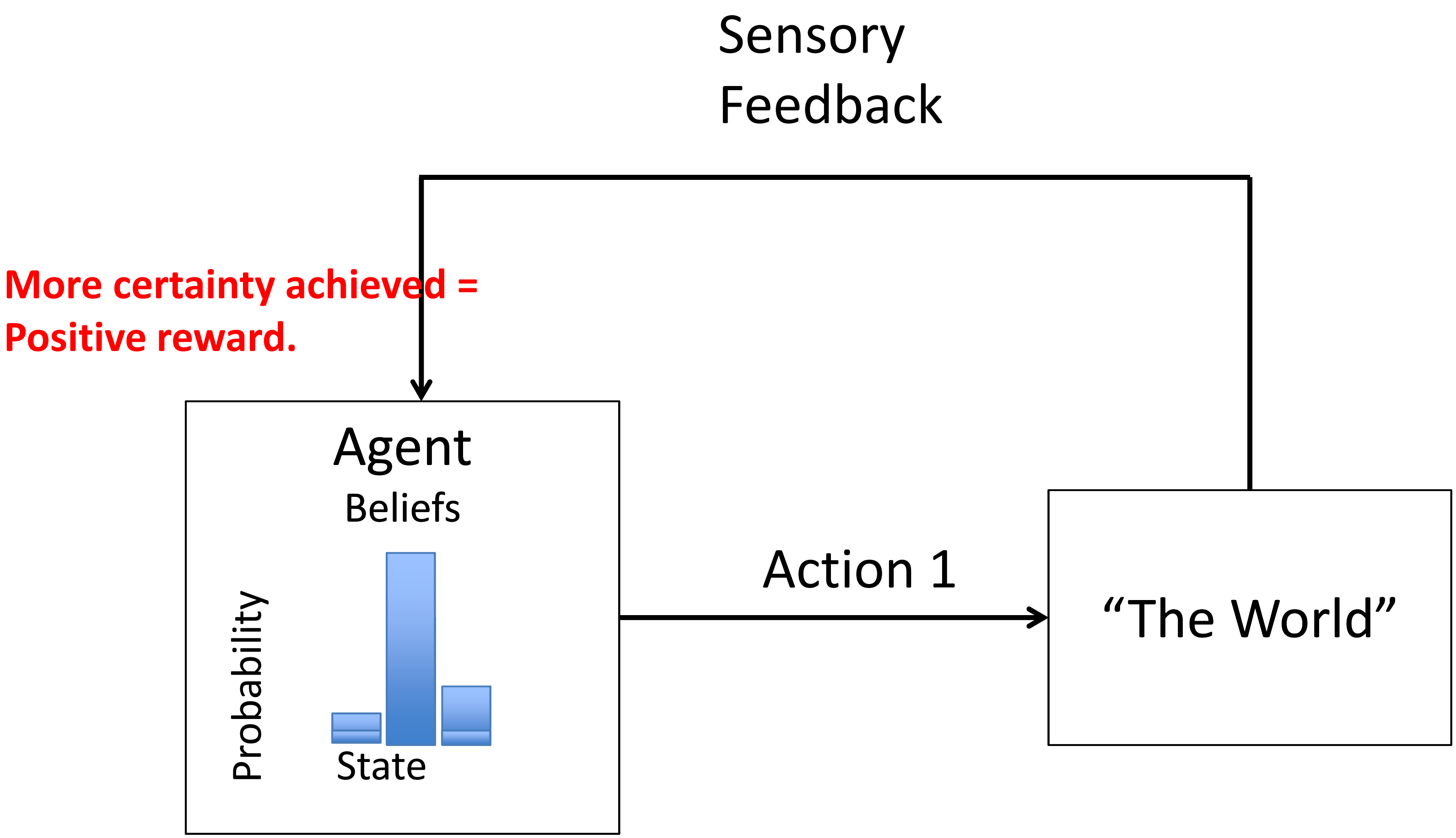
A (Slightly) More Complex Organism

Rewards and punishments are seldom unambiguous ("every cloud has its silver lining")

Decisions must be made in the face of highly uncertain and delayed rewards.

In order to be successful, organisms (and robots) must evaluate the efficacy of their own behavior.

Information is its Own Reward: Schematic of an InfoMax Agent

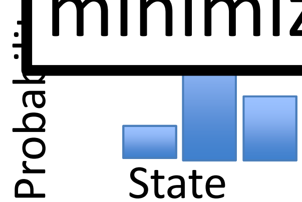


Information is its Own Reward: Schematic of an InfoMax Agent

The agent's reward signal is generated internally, rather than given externally by world.

No change in certainty = no reward

InfoMax agents choose actions that minimize uncertainty about the world.



POMDPs

POMDP: Partially Observable Markov Decision Process

Objective: choose actions to control a stochastic system optimally

$$\text{optimal behavior} = \underset{behavior}{\arg\max}\, E\left[\sum_{t=0}^{\infty} \gamma^t r(X_t) \mid \text{behavior}\right]$$

System's state, X_t, governed by stochastic dynamics

$$p(X_t | X_{t-1}, A_{t-1}) \longleftarrow \textbf{e.g. Newtonian Physics}$$

Information about the system's state must be inferred using noisy observations

$$p(O_t | X_t) \longleftarrow \textbf{e.g. Sensory System}$$

POMDPs for InfoMax Control

All **the information needed** to make optimal decisions is **given by the belief state**

For discrete systems, beliefs written as:

$$b_{t,i} = p(X_t = i | o_1 \ldots o_t, a_1 \ldots a_t)$$

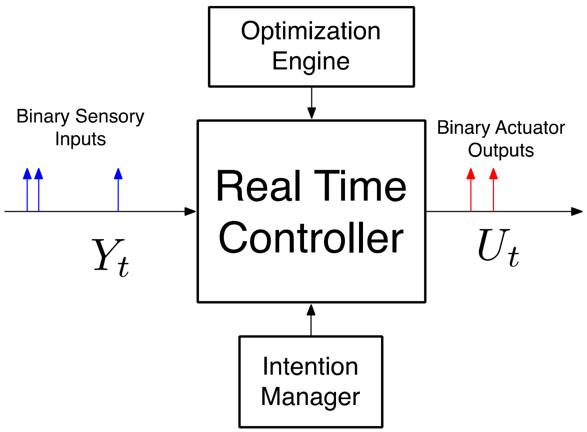
Reward value is the certainty of the belief state

$$r(b_t) = -H[b_t] = \sum_{i=1}^{n} b_{t,i} \log b_{t,i}$$

Example 1: Contingency Detection

Movellan and Watson (1985)

Computational Model



How should an infant probe the environment in an optimal fashion?

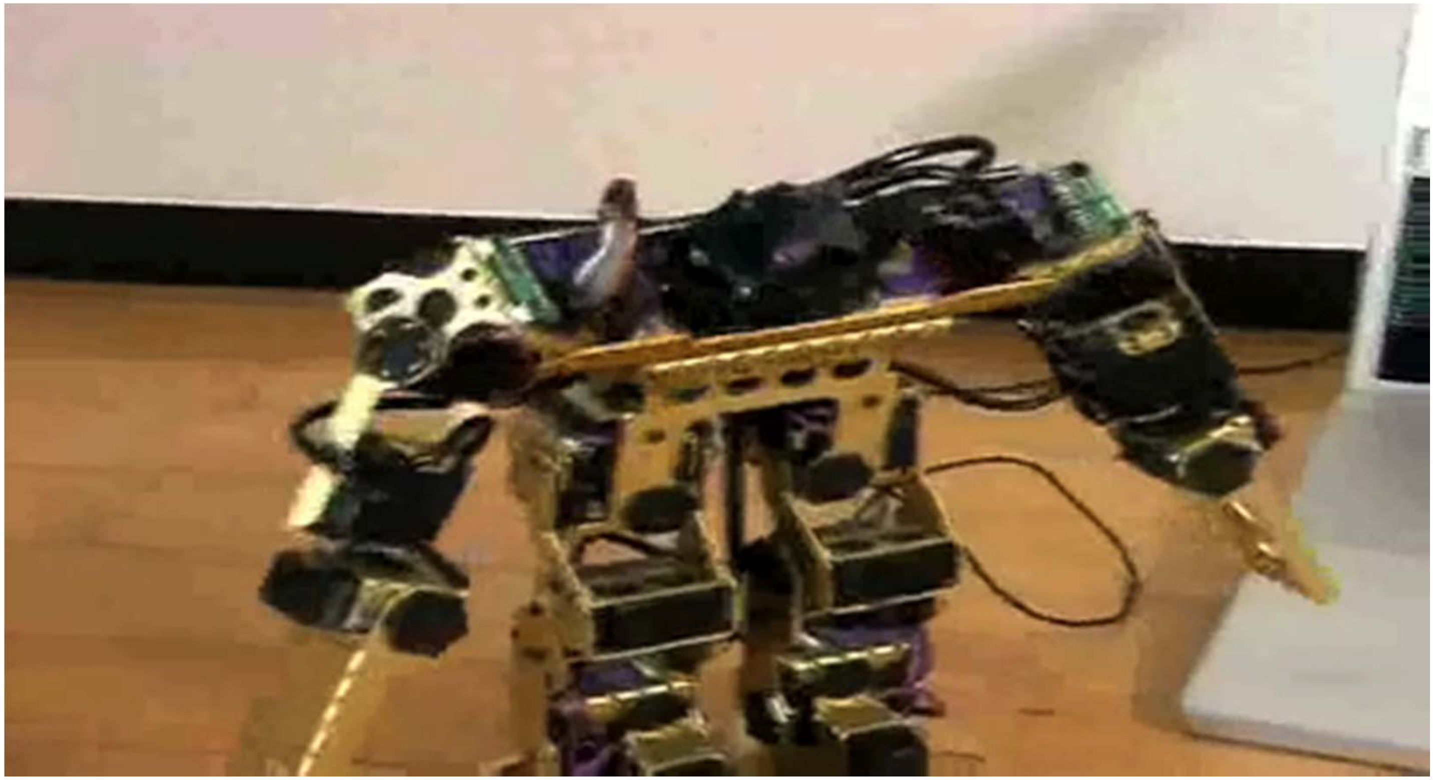
Mapping to POMDP:

- **State:** contingent caregiver present or not present
- **Actions:** vocalize or do not vocalize
- **Observations:** sound volume level
- **Observation model:** fit to data

Optimal behavior "emerges" from this problem description

Movellan (2005)

Contingency Detecting Robot



Example 2: Visual Search

We move our eyes 100,000 times a day. What is driving our choice of where to move them?

Theory: in part we move our eyes to maximize information about the state of the world

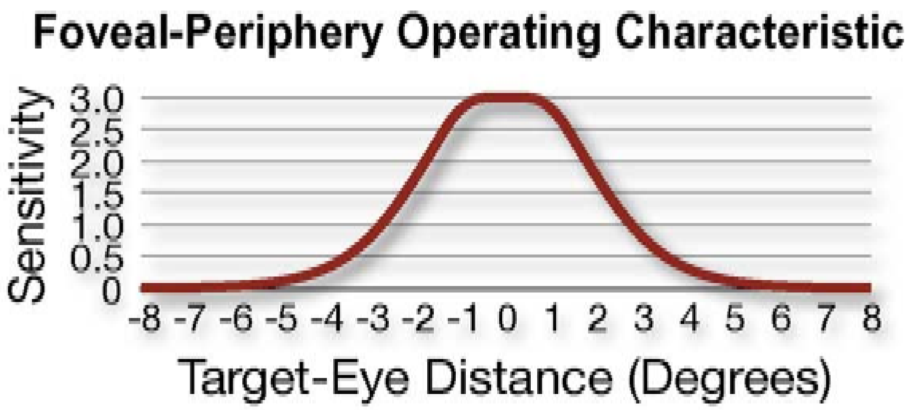
Mapping to POMDP:

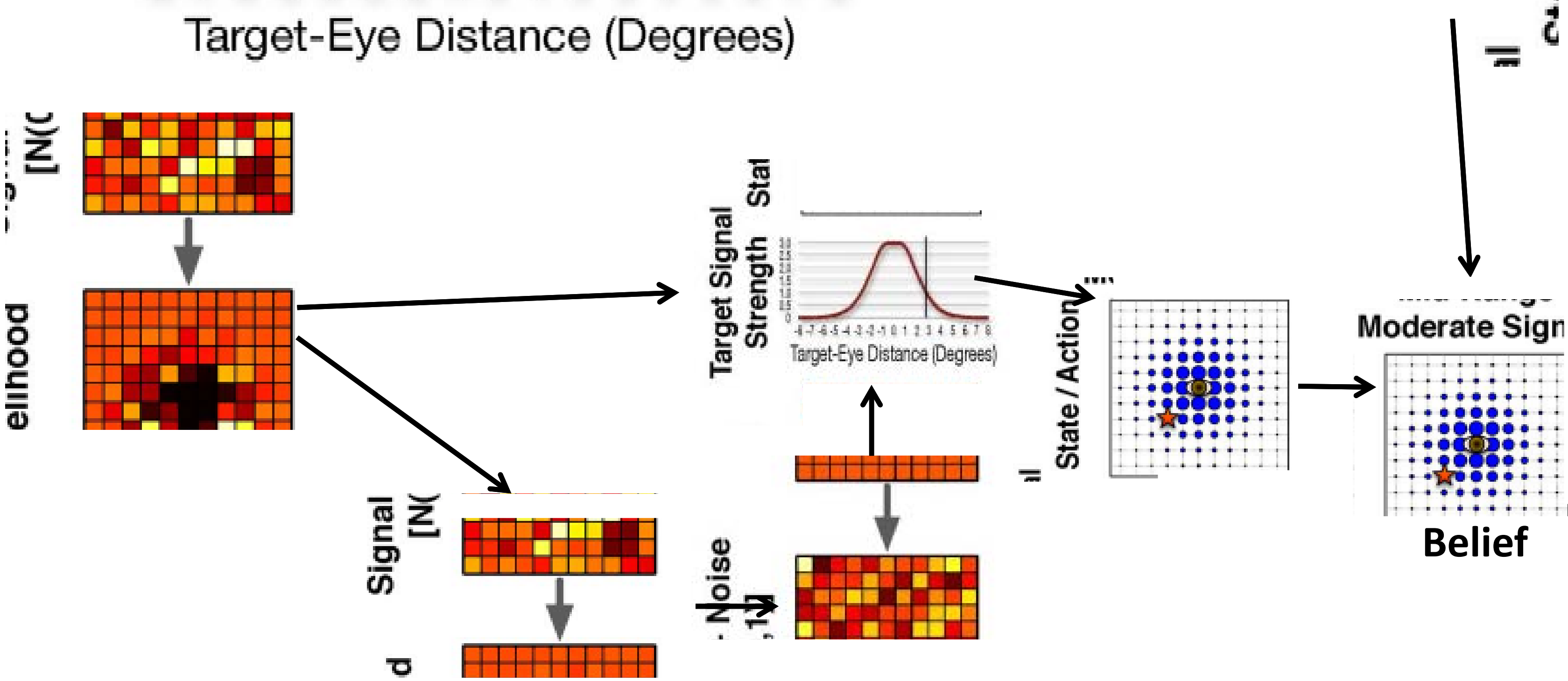
- **State:** location of some target of interest
- **Actions:** fixation position
- **Observations:** visual feedback
- **Observation model:** derived from psychophysical experiments

Bukto and Movellan (2008)

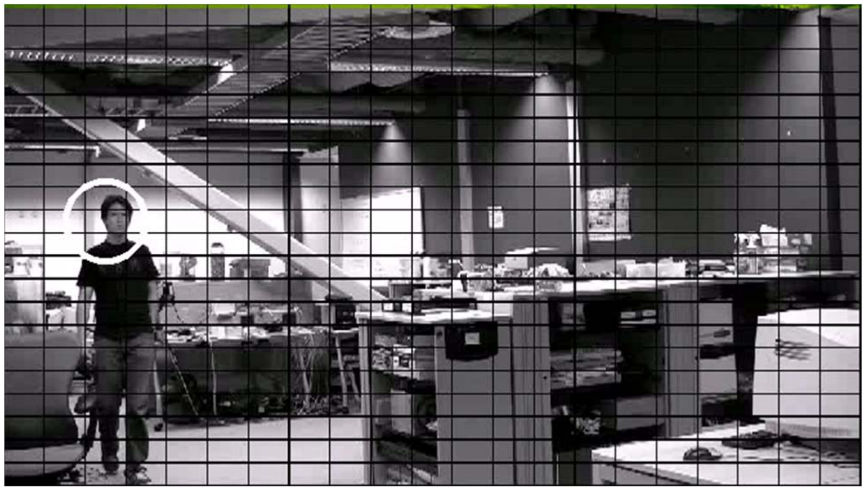
Extension: Fasel, Ruvolo, Wu, and Movellan (2009).

Sensory Model



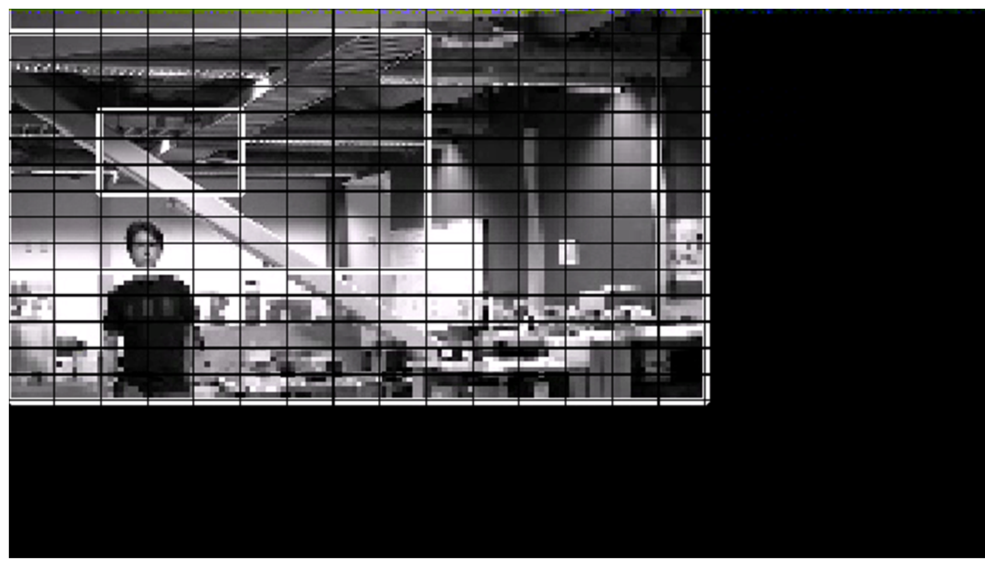


Digital Retina



Exhaustive Search

Foveated Search



Summary

Infomax Control is an example of a computational bridge to phenomena that students may not think of as computational.