Computational Psychiatry Course – Zürich 2018

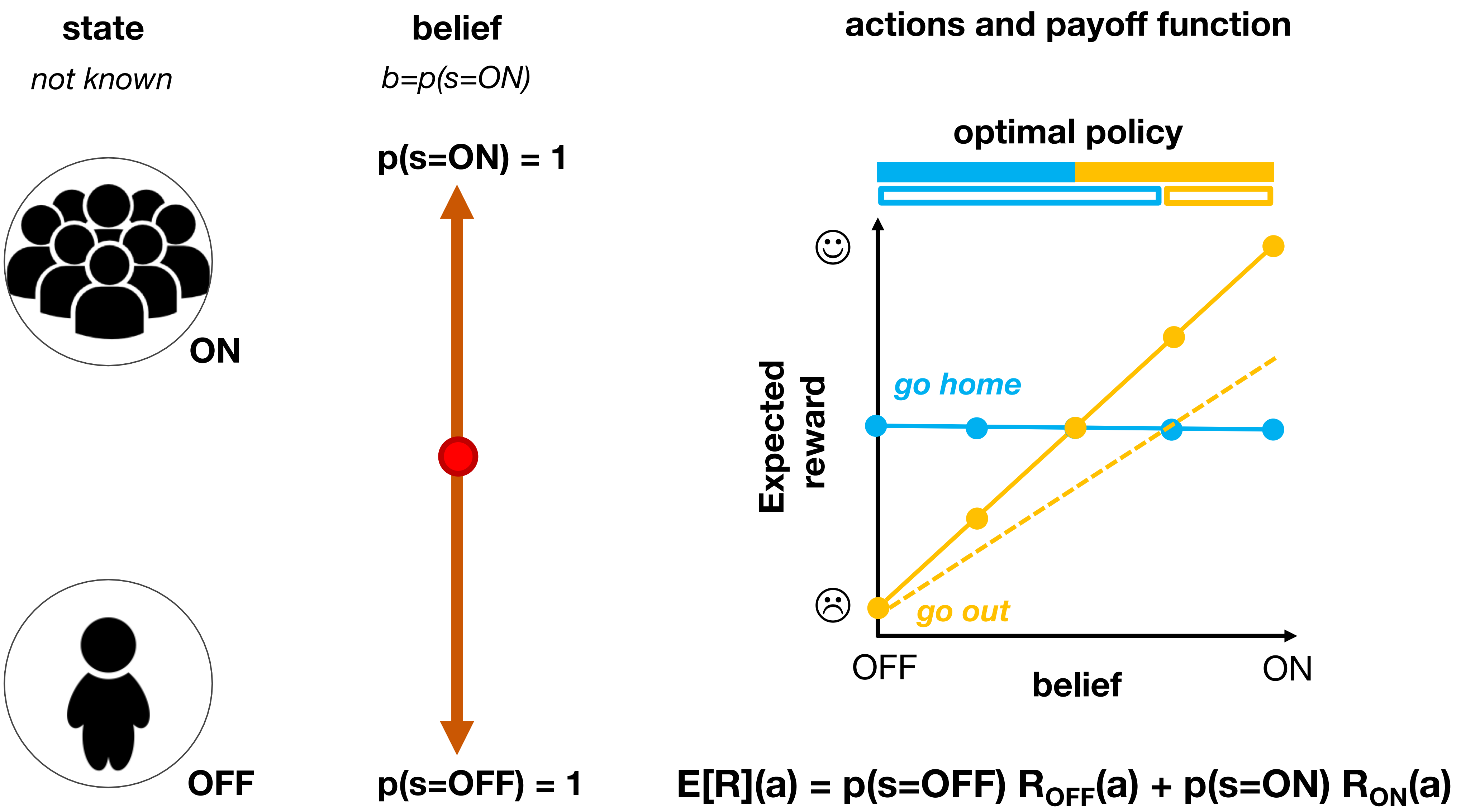
Partially Observable Markov Decision Processes (POMDP)

Lionel Rigoux & Frederike Petzschner

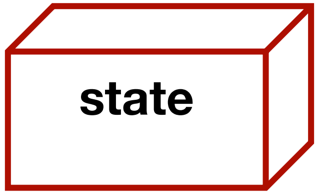
actions

go home		go out
	<	
	>	

go out

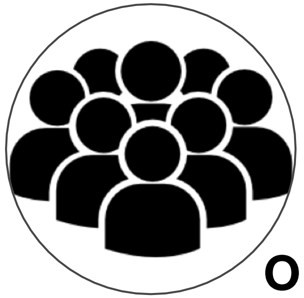


state
not known
ON
OFF
belief
b=p(s=ON)
p(s=ON) = 1
p(s=OFF) = 1
actions and payoff function
optimal policy
Expected reward
go home
go out
OFF
belief
ON
E[R](a) = p(s=OFF) R_{OFF}(a) + p(s=ON) R_{ON}(a)



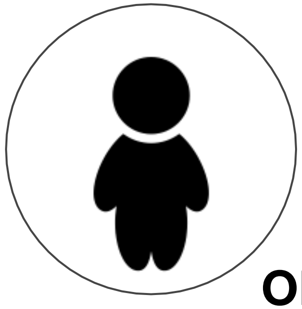
observation function

provide information about state



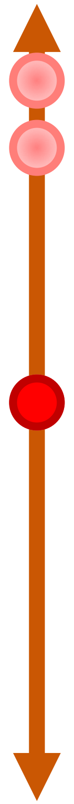
	phone	go out	go home
yes	0.85	1	0.5
no	0.15	0	0.5

$$b(s') \propto p(o|s', a) \sum_{s} p(s'|s, a) b(s)$$



	phone	go out	go home
yes	0.20	0	0.5
no	0.80	1	0.5

p(s=ON) = 1



p(s=OFF) = 1

POMDP Formalism

MDP

- S set of states
- A set of actions
- T transition matrix $S \times A \rightarrow S$
- R reward function $S \times A \rightarrow \mathbb{R}$
- γ discount factor

POMDP extension

- B belief space
- Ω set of observations
- O observation probabilities $S \times A \times \Omega \rightarrow \mathbb{R}^+$
- τ belief update function $B \times A \times \Omega \rightarrow B$ (Bayes)
- r reward function $B \times A \rightarrow \mathbb{R}$ (Expectation)

Optimal policy

$$\pi^\star = \arg\max_\pi \sum_{t=0}^{\infty} \gamma^t r(b_t, a_t)$$

Simulation workflow

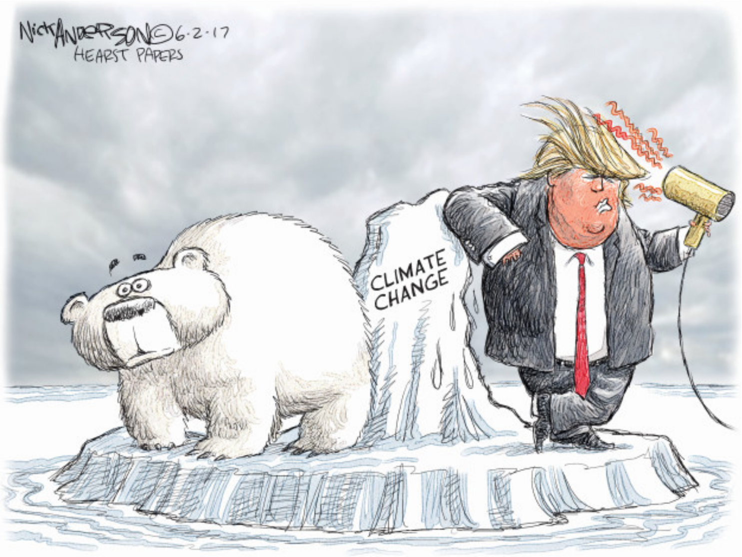
Initial state (s, b)

- Select action $a = \pi(b)$
- Update state $s' = T(s, a)$
- Receive outcome $R(s', a)$
- Get observation o $= O(s', a)$
- Update belief $b' = \tau(b, a, o)$
- Start over

Key properties

- **The value of information**

Information is valuable because (if) it improves control and therefore increases future outcomes



- **Subjective optimality**

Covert belief dynamics can explain seemingly economically irrational behavior

Take home message

POMDPs allow to model:

- sequential decision making in a changing environment (MDP)
- subjectivity (uncertainty) about the state of the world (PO)

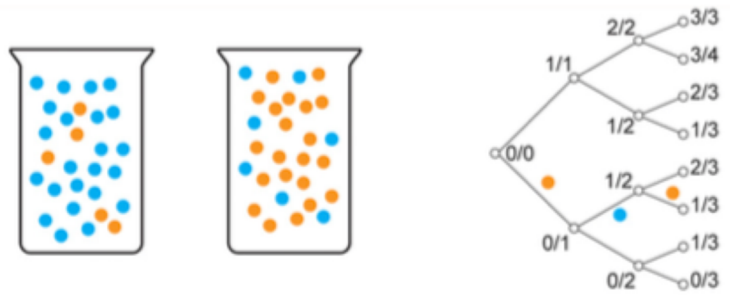
POMDPs can capture:

- information gathering as an economic decision
- irrational decision making as an optimal policy based on wrong representations

When to use?

>> Mix information accumulation with value optimization

- More complex belief updating
  - Hierarchical Bayesian Inference
  - Drift-Diffusion Model
- More complex value dynamics
  - Reinforcement learning
  - Classical MDP
- More complex belief updating & value dynamics
  - Ad hoc POMDP
  - Active inference

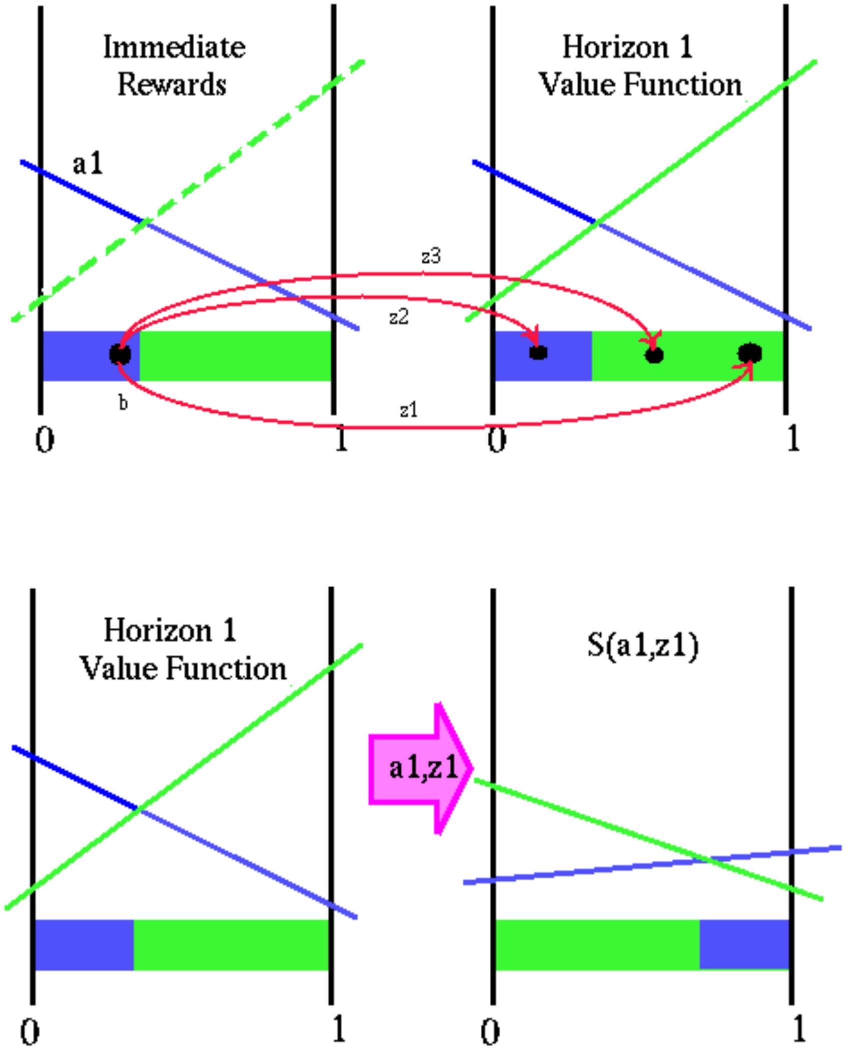


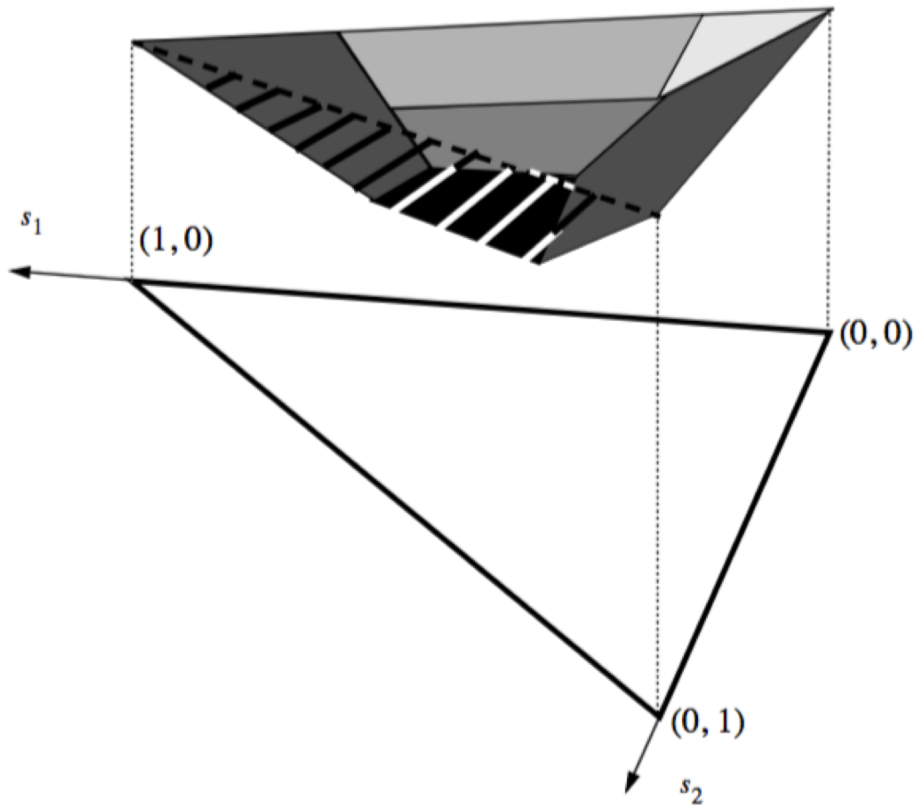
Averbeck 2015, PLoS CB
Choice in bandit task
Information sampling
Foraging

Paulus & Yu 2012, TICS
Emotions and decision

Questions?

Solution


Immediate
Rewards
a1
z3
z2
z1
b
0
1
Horizon 1
Value Function
0
1
Horizon 1
Value Function
a1,z1
S(a1,z1)
0
1
0
1


s_1
(1, 0)
(0, 0)
(0, 1)
s_2