Towards a rational theory of meta-decision making

TBD

Abstract

Keywords: Decision-Making; Heuristics; Meta Decision-Making

Introduction

big picture : decision strategies, heuristics, adaptive flexibility, and the debate about human rationality

our approach: resource-rationality, rational meta-decision making as the optimal solution to a meta-level MDP

payoffs:

1. a better normative standard of rational decision making that takes into account that people's time is finite and that their computational resources are bounded
2. an automatic way to discover novel decision strategies
3. new insights into how people make decisions under limited resources
4. an alternative to toolbox theories of judgment and decision making
5. a fairer judgment of human rationality

our specific contribution: a resource-rational model of decision-making in the Mouselab paradigm, empirical test of novel predictions, discovery of a new decision strategy

plan for th paper:

Markov Decision Processes

Each sequential decision problem can be modeled as a *Markov Decision Process* (MDP)

$$M = (\mathcal{S}, \mathcal{A}, T, \gamma, r, P_0), \tag{1}$$

where $\mathcal{S}$ is the set of states, $\mathcal{A}$ is the set of actions, $T(s,a,s')$ is the probability that the agent will transition from state s to state s' if it takes action a, $0 \leq \gamma \leq 1$ is the discount factor, $r(s,a,s')$ is the reward generated by this transition, and P_0 is the probability distribution of the initial state S_0 (Sutton & Barto, 1998). A *policy* $\pi : \mathcal{S} \mapsto \mathcal{A}$ specifies which action to take in each of the states. The expected sum of discounted rewards that a policy π will generate in the MDP M starting from a state s is known as its *value function*

$$V_M^{\pi}(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \cdot r(S_t, \pi(S_t), S_{t+1})\right]. \tag{2}$$

The optimal policy $\pi_M^{\star}$ maximizes the expected sum of discounted rewards, that is

$$\pi_M^{\star} = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \cdot r(S_t, \pi(S_t), S_{t+1})\right], \tag{3}$$

Deciding how to decide

The adaptive decision-maker

1. The Mouselab paradigm: many alternative decision strategies
2. information acquisitions as a window on the decision process

Optimal meta-decision-making

1. Meta-level MDPs as a computational-level theory of deciding how to decide (Hay, Russell, Tolpin, & Shimony, 2012).
2. Meta-level MDP of the Mouselab task
3. Approximating the optimal meta-level policy: Bayesian value function approximation

Experimental Test of novel predictions

Model Predictions

1. emergence of familiar decision strategies like TTB and WADD for specific problems
2. problem-contingent "strategy selection" including the effect of compensatoriness
3. previously unobserved effects of people's prior knowledge about the distribution of possible payoffs on their decision process
4. Previously unobserved SAT-TTB hybrid strategy terminates decision process early when a high payoff is observed on a probable outcome and the range of payoffs is small compared to the cost of time
5. Information acquisition become systematically more frugal as the range of possible payoffs decreases

Methods

Participants: We will recruit 200 participants on Amazon Mechanical Turk. Based on ? (?), we expect the task to take about 30 minutes. Participants will receive a baseline payment of \$1.50 to guarantee a minimum rate of \$3 per hour, and can earn a bonus of up to \$9.99.

Procedure: Mouselab experiment with

- 2 blocks a 20 trials with 4 outcomes
- inspected outcomes remain visible on the screen
- no time limit
- participants receive the payoff from a randomly selected trial

Experimental Design: 2x2x2 within subjects design:

1. IV1: range of payoffs: manipulated within subjects across blocks either [$0.00; $0.25] vs. [$0.01; $9.99].
2. IV2: number of gambles: 2 vs. 7; 5 instances of each in each block
3. IV3: compensatoriness: highly non-compensatory (e.g., $[0.9, 0.05, 0.03, 0.02]$) vs. nearly uniform (e.g., $[0.35, 0.25, 0.2, 0.15]$); 5 instances of each in each block

Results

Discussion

1. summary, implications, and future directions
2. conclusion

Acknowledgments. This work was supported by grant number ONR MURI N00014-13-1-0341.

References

Hay, N., Russell, S., Tolpin, D., & Shimony, S. (2012). Selecting computations: Theory and applications. In N. de Freitas & K. Murphy (Eds.), *Uncertainty in artificial intelligence: Proceedings of the twenty-eighth conference.* P.O. Box 866 Corvallis, Oregon 97339 USA: AUAI Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA, USA: MIT press.