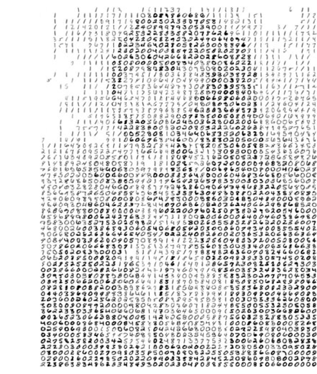
# Optimal nudging

Mathew D. Hardy, Frederick C. Callaway, & Thomas L. Griffiths
**Princeton University**
mdhardy@princeton.edu, fredcallaway@princeton.edu tomg@princeton.edu



## Background

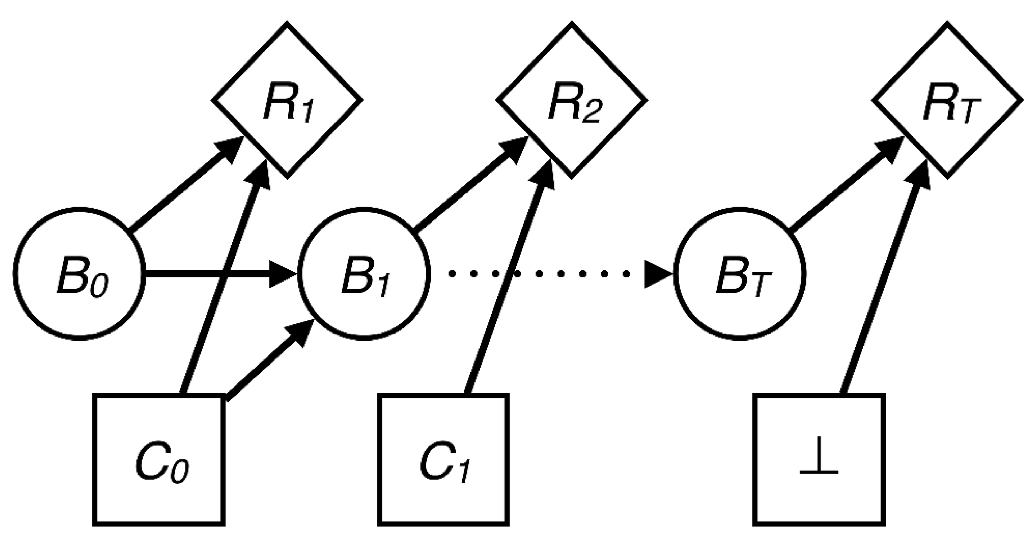
- People often make costly errors in their decisions [1].
- To mitigate these individual and social costs, researchers have developed nudge theory as a way to lead people to better options without restricting their freedom of choice [2].
- However, there is disagreement about the desired effects of nudges, and their effects on choice can be hard to predict.

## Research goals

- We propose a method for predicting and explaining the effects of choice architecture on option selection.
- We show how this approach allows flexible construction of optimal nudges, or those that maximize some objective function.

## Metalevel Markov decision processes

- We model an individual's deliberation process as a metalevel Markov decision process (metalevel MDP), which treats reasoning as a sequential decision problem [3].
- A metalevel MDP is defined analogously to a standard MDP, $(\mathcal{B}, \mathcal{C}, T_{\text{meta}}, r_{\text{meta}})$, where states $\mathcal{B}$ correspond to beliefs, and actions $\mathcal{C}$ correspond to cognitive operations, or computations.




**References**
[1] Kahneman, D., Slovic, P., & Tversky, A. e. (1982). Judgment under uncertainty: heuristics and biases. Cambridge, England: Cambridge University Press.
[2] Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. New Haven: Yale University Press.
[3] Hay, N., Russell, S., Tolpin, D., & Shimony, S. E. (2012). Selecting computations: Theory and applications. In Proceedings of the 28th conference on uncertainty in artificial intelligence.
[4]. Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(3), 534–552.

## Modeling optimal nudging

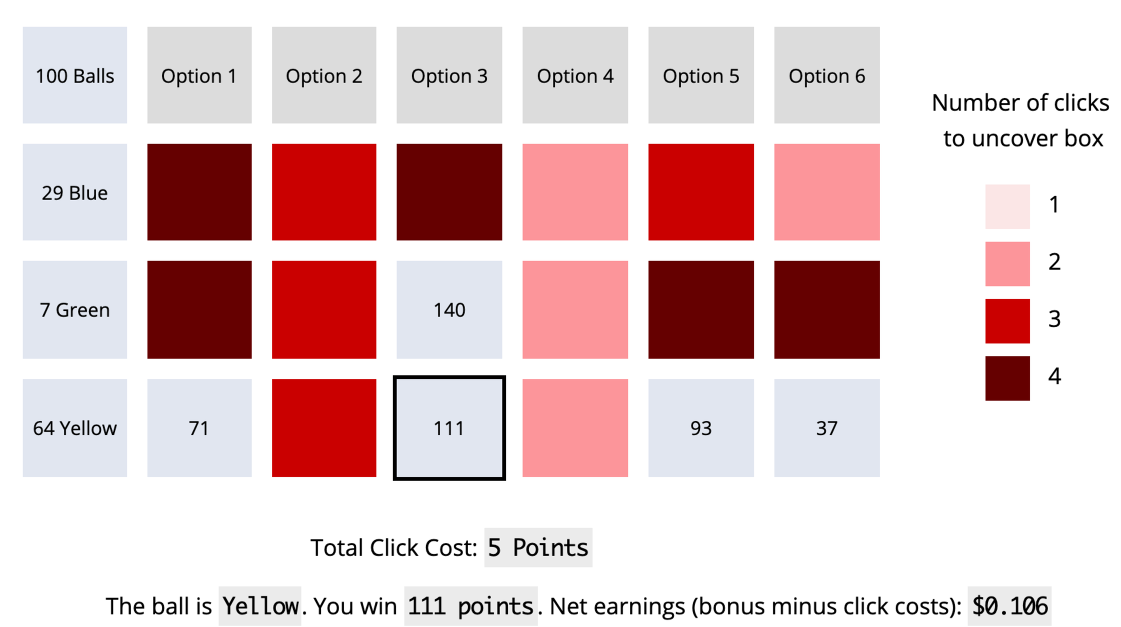
- We model nudges as modifications of the costs of cognitive computations available to the decision maker.
- These changes in costs can change both the sequence of computations and the eventual decision of the decision maker.
- An optimal nudge is thus defined as the cost modification that maximizes some function of the decision maker's deliberative actions and eventual choice.
- While many possible objectives are possible — and could in principle be chosen by the decision maker — we focus on the case where the optimal nudge is defined as the modification that maximizes metalevel reward:

$$\mathbb{E}\left[\sum_{t}^{T} r(B_t, C_t) \middle| \tilde{\lambda}, \theta^*, \pi_{\text{meta}}\right].$$

- Here $\tilde{\lambda}$ denotes the modified costs of $\mathcal{C}$, $\theta^*$ the true values of the decision-relevant parameters that are being estimated, and $\pi_{\text{meta}}$ the metalevel policy of the decision maker.

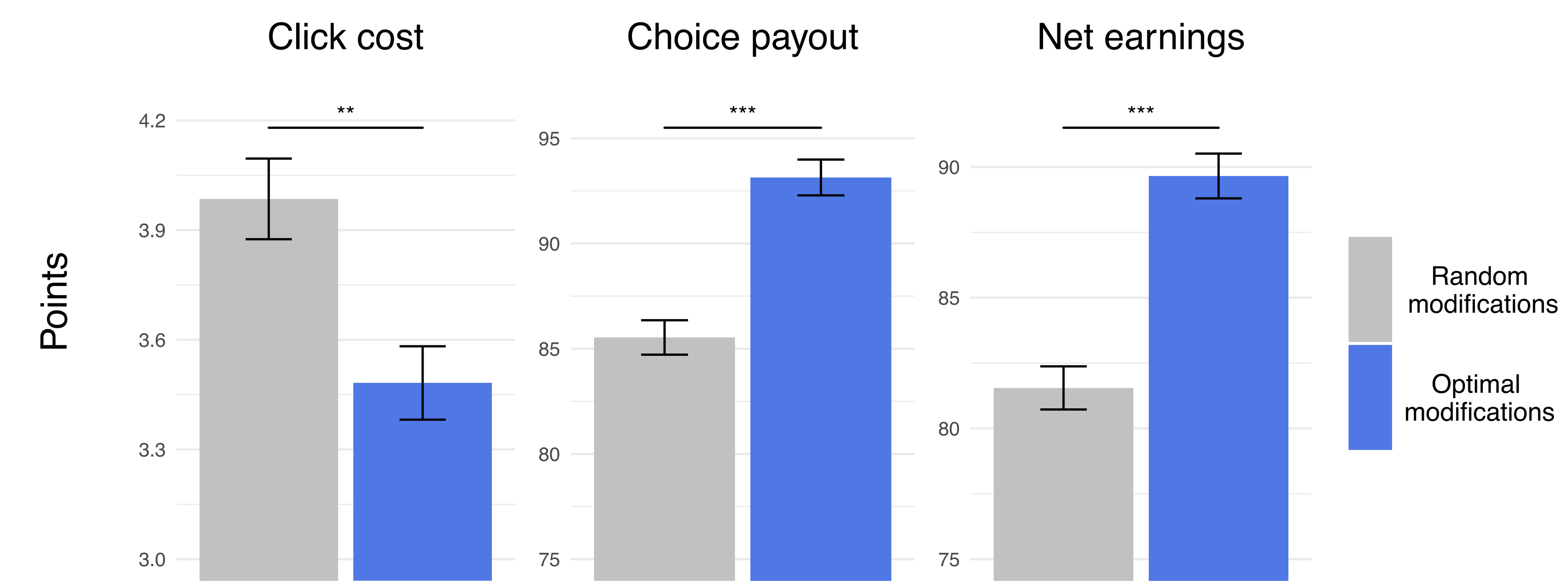
## Experiment

- We tested our proposed optimal nudging framework to a modified version of the Mouselab paradigm [4].
- In this setup, computations are externalized as costly information-gathering clicks that inform a multialternative choice.




## Results

- 150 participants were recruited from Amazon Mechanical Turk and made decisions on 20 problems.
- For each problem, the cost of uncovering a value was randomly determined, after which either a random or optimal modification was added to the problem.
- Optimal modifications were those that maximized the expected metalevel reward (choice payout minus click cost) of the metagreedy policy, and random modifications reduced the costs of three random values by an equal total amount.




## Conclusions

- Many types of nudges can be modeled as modifications to the computational costs of attaining certain pieces of information.
- Optimal cost modifications (i.e., nudges) both increased choice payout and decreased click cost compared to random modifcations.
- Our approach allows the optimization of arbitrary objective functions which individuals could in principle select.
- In future work, we hope to apply this framework to more everyday decisions in which computational costs are not externalized and have to be inferred from behavior.