Discovering Rational Heuristics for Risky Choice

Sayan Gul[1,*], Paul M. Krueger[1,*], Frederick Callaway[1,*], Thomas L. Griffiths[1], Falk Lieder[1,2,*]
[1] UC Berkeley, [2] MPI for Intelligent Systems, Tübingen [*] Equal contribution.

Abstract

How should we think and decide to make the best possible use of our precious time and limited cognitive resources? And how do people's cognitive strategies compare to this ideal? We study these questions in the domain of multi-alternative risky choice using the methodology of resource-rational analysis. To answer the first question, we leverage a new meta-level reinforcement learning algorithm to derive optimal heuristics for four different risky choice environments. We find that our method rediscovers two fast-and-frugal heuristics that people are known to use, namely Take-The-Best and choosing randomly, as resource-rational strategies for specific environments. Our method also discovered a novel heuristic that combines elements of Take-The-Best and Satisficing. To answer the second question, we use the Mouselab paradigm to measure how people's decision strategies compare to the predictions of our resource-rational analysis. We found that our resource-rational analysis correctly predicted which strategies people use and under which conditions they use them. While people generally tend to make rational use of their limited resources overall, their strategy choices do not always fully exploit the structure of each decision problem. Overall, people's decision operations were about 88% as resource-rational as they could possibly be. A formal model comparison confirmed that our resource-rational model explained people's decision strategies significantly better than the Directed Cognition model of Gabaix et al. (2006). Our study is a proof-of-concept that optimal cognitive strategies can be automatically derived from the principle of resource-rationality. Our results suggest that resource-rational analysis is a promising approach for uncovering people's cognitive strategies and revisiting the debate about human rationality with a more realistic normative standard.

Introduction and background Making good decisions is a very challenging computational problem. People have to make tens of thousands of decisions every day. This necessitates very efficient decision strategies, but there is currently no systematic way to discover such strategies and prove that they are optimal. Here, we introduce a novel computational method for discovering optimal heuristics for choosing between multiple risky alternatives and evaluate its predictions against human performance and the Directed Cognition model (1) according to which people's strategies are a myopic approximation to optimal decision-making.

People's decision strategies for multi-alternative risky choice have been extensively studied in the Mouselab paradigm (2) where participants choose between multiple gambles whose payoffs depend on a random outcome. Participants are shown the probability of each outcome and a payoff matrix with one column for each gamble and one row for each outcome. The entry in the o^{th} cell of row g indicates how much money gamble g pays if outcome o occurs. Critically, all payoffs are initially occluded, and the player can reveal outcomes by clicking on them one-by-one. Thus, the sequence of clicks a player makes traces their decision strategy. Each click is costly. Thus, to maximize earnings, the player must employ a decision strategy that achieves an optimal trade-off between the cost of clicking versus the value of information.

Deriving resource-rational heuristics for risky choice Uncovering people's decision strategies is a subject of ongoing research, and it remains unclear to which extent and under which conditions people's heuristics are rational. We address these problems by deriving resource-rational heuristics for multi-alternative risky choice. To do so, we formalized the problem of deciding how to decide as a meta-level MDP (3) and approximated its optimal solution using our recently developed Bayesian meta-level policy search method (4).

To validate this method on the Mouselab paradigm, we tested it on a small version of this task (3 alternatives, 2 outcomes, 8 payoffs) for which optimal strategies can be computed exactly. The discovered strategies achieved at least 99% of optimal decision-making across the four decision environments of the experiment reported below. Given these encouraging results, we proceeded to apply our method to a larger version of this paradigm for which the exact solution is no longer tractable.

Concretely, we applied this method to choices with 7 alternatives and 4 possible outcomes. We studied how the optimal heuristic depends on the stakes of the decision (low-stakes: \$0.01–\$0.25, high-stakes: \$0.01–\$9.99) and on whether one of its outcomes is much more probable than all others (*high dispersion*: $\max_o \Pr(o) \geq 0.85$ versus *low dispersion*: $0.15 \leq \Pr(o) \leq 0.40$ for all o). The cost of computation was always set to 1 cent per click.

We found that our method rediscovered the Take-The-Best heuristic (TTB) as the resource-rational strategy for high-stakes environments where one outcome is much more likely than all others, and it rediscovered choosing randomly (Random) as the resource-rational heuristic for problems with low-stakes where all outcomes are almost equally probable. Furthermore, our method discovered a previously unknown heuristic that combines elements of Take-The-Best and Satisficing (SAT-TTB). Like TTB, SAT-TTB inspects only the payoffs for the most probable outcome. But unlike TTB and like satisficing, SAT-TTB terminates as soon as it finds a gamble whose payoff for the most probable outcome is high enough. According to our method, SAT-TTB is the resource-rational heuristic for low-stakes problems where all outcomes are almost equally probable. For high-stakes problems with low-dispersion the resource-rational strategy often started with TTB but then inspected additional outcomes afterward (TTB+), or it inspected a subset of the two most probable outcomes (SAT-TTB2) or the three most probable outcomes (SAT-TTB3). Figure 1 shows a detailed summary of our model's predictions and compares them to the predictions of the Directed Cognition model and human data.

Experimental test of novel predictions To test the predictions of our resource-rational model, we designed an experiment presenting people with the four types of decision problems simulated above. We recruited 100 participants on Amazon Mechanical Turk, and paid them $\$0.25$ for about 13.5 min of work plus a performance-dependent bonus of up to $\$5.12$ (average bonus $\$3.39$). The experiment used a 2×2 within-subjects design. It was structured into two blocks of 10 choices with 7 alternatives and 4 outcomes each. In one block, all decisions had low-stakes ($\$0.01$–$\$0.25$); in the other block all decisions had high-stakes ($\$0.01$–$\$9.99$).Within each block, the outcome probabilities had low dispersion for half of the trials and high dispersion for the other half. These decision problems were presented using a variant of the Mouselab paradigm (2) where a) participants were charged $\$0.01$ per click, b) all inspected payoffs remained visible throughout the trial, and c) participants had to spend at least 30 sec on each trial before they were allowed to move on. After the last trial, 1 high-stake trial and 1 low-stake trials were randomly selected and each participant received the average of their net earnings on those two trials as a bonus.

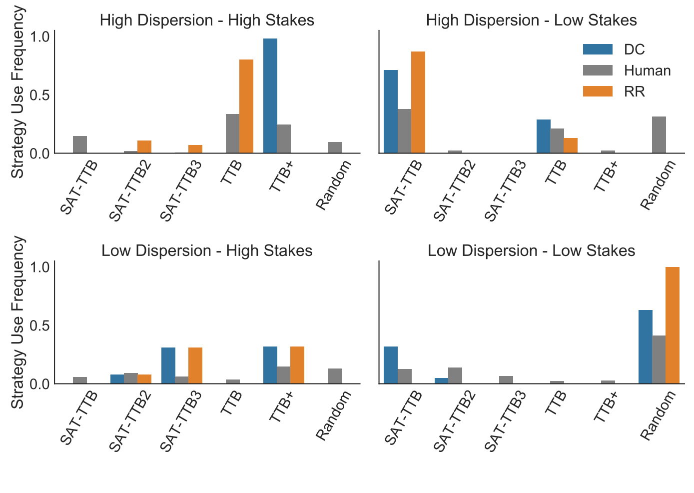

Figure 1: Strategy use frequencies of people, our resource-rational model, and the Directed Cognition model.

We found that participants did indeed use the previously unnoticed SAT-TTB heuristic discovered by our method; in fact, people used SAT-TTB more frequently than any other heuristic; just as our resource-rational analysis had predicted (see Figure 1). Furthermore, our resource-rational analysis accurately predicted how the frequency with which people use each heuristic depends on the stakes of the decision and the dispersion and the uncertainty about its outcome. Consistent with our model's predictions, participants used the fast-and-frugal heuristics TTB and SAT-TTB more frequently when one outcome was much more probable than all other outcomes than when all outcomes were almost equally probable (39% vs. 74%; $\chi^2(1) = 21.86, p =< 0.0001$). Furthermore, as the stakes increased people switched to more effortful and more accurate strategies. When the stakes were high, people considered a larger number of possible outcomes (9.8 ± 7.9 vs. 2.9 ± 3.1, $t(1998) = 25.4, p < 0.0001$) and used simple heuristics less often. Overall, the frequency with which people relied on fast-and-frugal heuristics (TTB, SAT-TTB, SAT, or random choice) decreased significantly from 73.1% on low-stakes problems to 40.1% on high-stakes problems ($\chi^2(1) = 116.9, p < 0.0001$). These findings suggest that people adapt their strategy use to the structure of the environment in accordance with the predictions of our resource-rational model. However, as Figure 1 shows, people's strategy choices did not change radically enough with the stakes and dispersion to be completely resource-rational.

Consistent with the idea that people first choose a decision strategy and then execute it, we found that participants deliberated longer before the first click (5.2 sec) than before subsequent clicks (3.1 sec, $t(14087) = 14.4, p < 0.0001$). Interestingly, they also deliberated more when the stakes were low (5.3 sec) than when the stakes were high (2.6 sec, $t(14087) = 14.4, p < 0.0001$). This may be because with low stakes the relative cost of unnecessary clicks is much higher.

To assess how resource-rational people's decision-strategies are, we evaluated their net-performance (payoff minus decision cost) against the net-performance of our resource-rational model. This revealed that, on average, the net-performance of our participants' decision-strategies was about 88.4% of the net-performance of resource-rational decision-making. The way in which people deviated from resource-rational decision-making depended on stakes and dispersion: When the stakes were low and the dispersion was high then collecting too little information was a more common mistake than collecting too much information (48.4% vs. 37.2%); but in all other conditions people tended to collect too much rather than too little information (52.3% vs. 10.0%). Furthermore, when the stakes were high, people's most common mistake was to inspect a payoff other than the one that the resource-rational strategy would have inspected (61.6% of all errors).

A formal model comparison showed that, overall, the resource-rational model explained participants' click sequences significantly better than the Directed Cognition model and a null model that selects computations at random ($\text{BIC}_{\text{RR}} = 79360.88 < \text{BIC}_{\text{DC}} = 81606.35 < \text{BIC}_{\text{random}} = 89172.66$; $\text{AIC}_{\text{RR}} = 78785.03 < \text{AIC}_{\text{DC}} = 81030.51 < \text{AIC}_{\text{random}} = 89172.66$). At the individual level, the data provided strong evidence for the resource-rational model over its alternatives for the majority of our participants: according to both the BIC and the AIC the click sequences of $64/100$ participants were best explained by the resource-rational model, compared to only $22/100$ for the Directed Cognition model, and $14/100$ for the null model.

Conclusion These findings are a proof-of-concept that optimal cognitive strategies can be automatically derived as the rational use of finite time and bounded cognitive resources. Our findings suggest that this is a promising approach to understanding how people think and decide that can cast new light on the debate about human rationality.

Acknowledgments This research was supported by a grant from the Templeton World Charity Foundation to T.L.G.

References

[1] X. Gabaix, D. Laibson, G. Moloche, S. Weinberg, *American Economic Review* **96**, 1043 (2006).
[2] J. W. Payne, J. R. Bettman, E. J. Johnson, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**, 534 (1988).
[3] F. Lieder, P. M. Krueger, T. L. Griffiths, *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, G. Gunzelmann, A. Howes, T. Tenbrink, E. J. Davelaar, eds. (Cognitive Science Society, Austin, TX, 2017), pp. 742–747.
[4] F. Lieder, F. Callaway, S. Gul, P. M. Krueger, T. L. Griffiths, *NIPS workshop on Cognitively Informed AI* (2017).