A resource-rational analysis of human planning

Frederick Callaway[1] (fredcallaway@berkeley.edu), Falk Lieder[1] (falk.lieder@berkeley.edu)

Priyam Das (priyam@berkeley.edu), Sayan Gul (sayangul@berkeley.edu),

Paul M. Krueger (pmk@berkeley.edu), Thomas L. Griffiths (tom_griffiths@berkeley.edu)

Department of Psychology, University of California, Berkeley

[1] These authors contributed equally.

Abstract

People's cognitive strategies are jointly shaped by function and computational constraints. Resource-rational analysis leverages these constraints to derive rational models of people's cognitive strategies from the assumption that people make rational use of limited cognitive resources. We present a resource-rational analysis of planning and evaluate its predictions in a newly developed process tracing paradigm. In Experiment 1, we find that a resource-rational planning strategy predicts the process by which people plan more accurately than previous models of planning. Furthermore, in Experiment 2, we find that it also captures how people's planning strategies adapt to the structure of the environment. In addition, our approach allows us to quantify for the first time how close people's planning strategies are to being resource-rational and to characterize in which ways they conform to and deviate from optimal planning.

Keywords: bounded rationality; planning; rational analysis; decision-making; heuristics

Introduction

Previous research has shown that many aspects of human cognition can be understood as rational adaptations to the environment and the goals people pursue in it (Anderson, 1990). *Rational analysis* leverages this assumption to derive models of human behavior from the structure of the environment. In doing so, rational analysis makes only minimal assumptions about cognitive constraints. However, it has been argued that there are many cases where the constraints imposed by cognitive limitations are rather substantial, and Herbert Simon famously argued that to understand people's cognitive strategies we have to consider both the structure of the environment and cognitive constraints simultaneously (Simon, 1956, 1982). *Resource-rational analysis* (Griffiths, Lieder, & Goodman, 2015) thus extends rational analysis to also take into account which cognitive operations are available to people, how long they take, and how costly they are. Given that resource-rational analysis has been successful at explaining a wide range of cognitive biases in judgment (Lieder, Griffiths, Huys, & Goodman, 2017) and decision-making (Lieder, Griffiths, & Hsu, 2017) by suggesting resource-efficient cognitive mechanisms, it might also be able to shed new light on other cognitive processes, such as planning.

Surprisingly little is known about how people plan. While extant models of planning (De Groot, 1965; Huys et al., 2012, 2015; Newell & Simon, 1956, 1972) explain aspects of human planning, its precise mechanisms remain unclear; the applicability of each existing model is limited; and it remains unknown when people use which of those strategies and why. These questions are very difficult to answer because planning is an unobservable and highly complex cognitive process.

Here, we address these problems by deriving planning strategies through resource-rational analysis and introducing a process-tracing paradigm that allows us to directly observe the sequence of people's planning operations. We use data obtained with this paradigm to quantitatively evaluate our resource-rational model of planning against previously proposed planning strategies. This enables us to discern between those models even when they predict the same final decision.

Our resource-rational framework enables us to automatically discover the optimal planning strategy for any given environment. We find that people's planning strategies are better explained by bounded-optimal planning than by classic models of planning as search (progressive deepening, best-first search, depth-first search, and breadth-first search) even when those models are augmented with the mechanisms of satisficing (Simon, 1956) and pruning (Huys et al., 2012). Our resource-rational analysis allows us to characterize how human planning conforms to and deviates from bounded-optimal planning and to quantify individual differences in the rationality of people's planning strategies. Finally, our analysis also correctly predicted how people's planning strategies differ across environments.

This paper is structured as follows. We start by introducing the methodology of resource-rational analysis and review previous findings on planning. Next, we introduce our new process-tracing paradigm for the study of planning and apply resource-rational analysis to its planning problems. We then evaluate the resource-rational model against process-tracing data from people in Experiment 1. Experiment 2 tests resource-rational predictions about how people's planning strategies should change with the structure of the environment. We close by discussing the implications of our findings for cognitive modeling and human rationality.

Background

Discovering optimal cognitive strategies

Resource-rational analysis (Griffiths et al., 2015) derives process models of how cognitive abilities are realized from a formal specification of their function and a model of the cognitive architecture available to realize them. Formally, the resource-rational model of a cognitive mechanism is defined as the solution to a constrained optimization problem over the space of strategies that can be implemented on the assumed cognitive architecture, and the objective function measures how well the strategy would perform under the constraints of limited time and costly computation.

Planning

Most research on planning has been conducted in the fields of problem solving and artificial intelligence (Newell & Simon, 1972). The Logic Theorist (Newell & Simon, 1956) planned its proofs using *breadth-first search*: it first evaluated all possible one-step plans, then proceeded to all possible two-step plans, and so on, until it discovered a proof. By contrast, chess programs typically use *depth-first search*: they evaluate one possible continuation in depth and then back up one step at a time. Newell and Simon's (1972) General Problem Solver planned backward by *means-ends analysis* which compares the current state to a goal state in order to identify actions that can be taken to reduce their discrepancy.

Newell and Simon's (1972) research on human problem solving found that people usually plan forwards by a strategy called *progressive deepening* (De Groot, 1965) which is similar to depth-first search but resumes planning from the beginning after having considered one action sequence in depth. Furthermore, Simon (1956) argued that human decision-making is fundamentally constrained by limited cognitive resources and that people cope with these constraints by choosing the first option they find good enough instead of trying to find the best option; this is known as *satisficing*.

More recent work has found that people often prune their decision tree when they encounter a large loss (Huys et al., 2012) and cache and reuse previous action sequences (Huys et al., 2015). Furthermore, it has been argued that people greedily choose each of their planning operations so as to maximize the immediate improvement in decision quality instead of considering the potential benefits of sequences of planning operations (Gabaix, Laibson, Moloche, & Weinberg, 2006).

The Mouselab-MDP paradigm

Planning, like all cognitive processes, cannot be observed directly. In previous work, researchers have inferred properties of human planning from the decisions participants ultimately made or asked participants to verbalize their planning process. However, many different planning strategies can lead to the same final decision, and introspective reports are often incomplete or inaccurate.

To address these challenges we employ a new *process-tracing paradigm* for the study of planning that externalizes people's unobservable beliefs and planning operations by observable states and actions (Callaway, Lieder, Krueger, & Griffiths, 2017). Inspired by the Mouselab paradigm (Payne, Bettman, & Johnson, 1993) that traces how people choose between multiple risky gambles, the Mouselab-MDP paradigm uses people's mouse-clicking as a window to their planning.

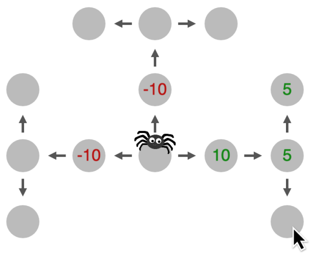

Figure 1: Illustration of the Mouselab-MDP paradigm. http://cocosci.dreamhosters.com/webexpt/webofcash-demo

Each trial presents a route planning problem where each location (the gray circles in Figure 1) harbors a reward or punishment. These potential gains and losses are initially occluded, corresponding to a highly uncertain belief state, but the participant can reveal each location's value by clicking on it and paying a fee. Clicking roughly corresponds to simulating or recalling the immediate reward of entering a potential future state and updating one's belief accordingly. Although this is likely a dramatic oversimplification of the representations and computations people employ when planning, it nevertheless retains enough of the core structure of planning to reveal previously unobservable aspects of human planning.

Bounded-optimal planning

Following the strategy discovery method by Lieder, Krueger, and Griffiths (2017), we model the optimal planning strategy for the Mouselab-MDP paradigm with uniformly distributed rewards as the solution to the meta-level Markov Decision Process (meta-level MDP; Hay et al., 2012)

$$M_{\text{meta}} = (\mathcal{B}, \mathcal{A}, \mathcal{T}, r_{\text{meta}}), \tag{1}$$

where each belief state $b \in \mathcal{B}$ encodes a uniform distribution over a move's reward. Thus, the belief state $b^{(t)}$ at time t can be represented as $\left(\mathcal{R}_1^{(t)}, \cdots, \mathcal{R}_K^{(t)}\right)$ where $\mathcal{R}_1^{(t)}$ is the set of possible values that the hidden reward X_k might take such that $b^{(t)}(X_k = x) = \text{Uniform}(x; \mathcal{R}_k^{(t)})$ and $b^{(0)}$ encodes the distribution the reward structure is sampled from. The meta-level actions are $\mathcal{A} = \{c_1, \cdots, c_K, \perp\}$ where c_k reveals the reward at state k and $\perp$ selects the path with highest expected sum of rewards according to the current belief state. The transition probabilities $T_{\text{meta}}(b^{(t)}, c_k, b^{(t+1)})$ encode that performing computation c_k sets $\mathcal{R}_k^{(t+1)}$ to $\{x\}$ with probability $1/\left|\mathcal{R}_k^{(t+1)}\right|$ for $x \in \mathcal{R}_k^{(t)}$. The meta-level reward function is $r_{\text{meta}}(b, c) = -\lambda$ for $c \in \{c_1, \cdots, c_K\}$, and

$$r_{\text{meta}}\left((\mathcal{R}_1, \cdots, \mathcal{R}_K), \perp\right) = \max_{\mathbf{t} \in \mathcal{T}} \sum_{k \in \mathbf{t}} \frac{1}{|\mathcal{R}_k|} \cdot \sum_{x \in \mathcal{R}_k} x, \tag{2}$$

where $\mathcal{T}$ is the set of all paths $\mathbf{t}$.

Having formulated the problem of deciding how to plan in these terms, we can now compute the optimal planning strategy by solving the meta-level MDP using backward induction (Puterman, 2014).

Experiment 1: Testing models of planning

In Experiment 1, we leveraged the Mouselab-MDP paradigm (Figure 1) to evaluate how people plan against bounded-optimal planning and classic models of planning.

Methods

Stimuli and Procedure The experiment started with a series of practice blocks (one for navigating with all rewards revealed, one for navigating with all rewards concealed, one for inspecting nodes, and a last one that introduced the cost of inspecting rewards) and a quiz that queried participants about the range of rewards, the cost per click, and their bonus.

In the main part of the experiment each participant solved 30 different 3-step planning problems of the form shown in Figure 1. There were 3 options for the first move and two options for the third move, leading to 6 paths in total. Each location's reward was independently drawn from a discrete uniform distribution over the values $-10, -5, +5$, and $+10$, and the cost of inspecting a node was $\lambda = \$1$. To reduce the opportunity cost of time, participants were required to spend at least 7 seconds on each trial. If they finished the trial in less time, then a countdown appeared and they were told to wait until the remaining seconds had passed.

Participants We recruited 60 participants from Amazon Mechanical Turk. Each participant received a base pay of \$0.50 and a performance-dependent bonus that was proportional to their score in the task (average bonus: \$2.16±\$1.16) for about 16.6 minutes of work on average. We excluded 9 participants (15%) because they either failed to follow the instruction to click during the training phase or answered fewer than 2 of the 3 comprehension checks correctly.

Models

We model people's planning operations c as arising from a combination of a systematic strategy M and unexplained variability according to

$$\Pr(c|b,M,\theta_M) = (1-\varepsilon)\cdot\sigma(c;V_{b,M},\tau) + \varepsilon\cdot\mathcal{U}(c;C_b), \quad (3)$$

where the first term models the strategy's choice of computations as a soft-max function ($\sigma(c;V_{b,M})$), and the second term models unexplained variability by a uniform distribution over the set C_b of all clicks that have not been made yet and the termination action $\perp$. The weight ε of the random process is a free parameter that we constrain to be less than 0.25. The probability $\sigma(c;V_{b,M})$ that the strategy will choose computation c is defined as a soft-max decision rule

$$\sigma(c;V_{b,M},\tau) = \frac{\exp(\frac{1}{\tau}\cdot V_{b,M}(c))}{\sum_{c'\in C_b}\exp(\frac{1}{\tau}\cdot V_{b,M}(c'))}, \quad (4)$$

over its preferences $V_{b,M}$. The decision temperature τ interpolates between always choosing the most preferred computation and choosing computations at random. The models presented below differ only in $V_{b,M}(c)$. In addition, we consider a null model that only includes the second component and thus chooses clicks uniformly at random.

Characterization of the bounded-optimal strategy We formalized the bounded-optimal planning strategy for Experiment 1 as the solution to a meta-level MDP M_{meta}. We computed the optimal meta-level Q-function $Q^{\star}_{\text{m}}$ of M_{meta} using backward induction. This allows us to define our model of bounded-optimal planning (M_{BO}) as

$$\Pr(c|b,M_{\text{BO}},(\tau,\varepsilon)) = (1-\varepsilon)\cdot\sigma(c;Q^{\star}_{\text{m}},\tau) + \varepsilon\cdot\mathcal{U}(c;C_b), \quad (5)$$

and the bounded-optimal planning strategy is given by

$$\Pr(c|b) = \lim_{\tau\to 0}\Pr(c|b,M_{\text{BO}},(\tau,0)). \quad (6)$$

We characterized this optimal strategy by inspecting the probability with which it chooses each planning operation across 40 simulated trials. We discovered that the bounded-optimal strategy is similar to best-first search in that the nodes it inspects always lie on one of the inspectable paths with the highest expected return. But it differs from best-first search in that it inspects those nodes in a different order and makes distinct predictions about when people should terminate planning. We test these qualitative predictions below.

Models of classical planning strategies To evaluate our bounded-optimal planning strategy against extant theories, we built likelihood models of the classical planning strategies known as depth-first search, breadth-first search, best-first search, and progressive deepening search (Newell & Simon, 1972). Each strategy M was defined by the values $V_{b,M}(c)$ it assigns to the different clicks $c \in C_b$ depending on the current belief state b (see Equation 4).The preference functions $V_{b,M}$ are defined such that $\sigma(c;V_{b,M},\tau = 10^{10})$ reproduces the behavior of the modeled strategy M. Depth-first search prioritizes deeper nodes on partially observed paths by setting $V_{b,\text{DFS}}(c)$ to the depth of the node inspected by c in the decision tree. Breadth-first search prioritizes shallower nodes on partially observed paths by setting $V_{b,\text{BFS}}(c)$ to minus the depth of the node inspected by computation c. Best-first search prioritizes nodes on promising paths by assigning the expected sum of rewards along the path on which c lies to $V_{b,\text{BestFS}}(c)$. Progressive deepening prioritizes deeper nodes adjacent to nodes that have been already inspected similarly to depth-first search. However, once a path is fully explored, the starting nodes of any sub-paths that may branch off from it receive the same value as the starting nodes of other, separate paths. For all strategies, paths are explored in the order they would be traversed.

Our models augment these classic search-based strategies with satisficing (Simon, 1956) and pruning (Huys et al., 2012): When the expected reward for terminating in the current state b equals or exceeds the model's aspiration level α, then $V_{b,M}(\perp) = 10^{10}$ so that all strategies M strongly prefer to terminate planning ($\perp$). If the expected return of a path falls below the pruning threshold ω, then $V_{b,M}(c) = -10^{20}$ for all computations c that inspect any of the nodes along that path; this ensures that none of those nodes will be inspected. Thus,

Model	BIC	AIC	LL
Optimal	30625	30611	-15303
Best First	31744	31716	-15854
Directed Cognition	34025	34011	-17004
Breadth First	34078	34058	-17026
Random	34579	34579	-17289
Progressive Deepening	34725	34704	-17349
Depth First	34732	34711	-17352

Table 1: Model comparison: Columns are Bayesian Information Criterion, Akaike Information Criterion, and Log Likelihood.

each of these models has two additional free parameters: the aspiration level α and the pruning threshold ω.

Directed cognition model We extended the directed cognition model (Gabaix et al., 2006) to the Mouselab-MDP paradigm. The directed cognition model uses macro-operators called options. For our task, each macro-operator is a sequence of clicks along a path. Therefore, each option can be defined by a (p, n_c)-tuple, where p is a path and n_c is the number of clicks the macro-operator makes along that path. Given an option, the nodes that are clicked along a path are picked in the order of decreasing variance, and the ties are broken at random. The directed cognition model chooses macro-operators according to a myopic cost-benefit analysis. Concretely, the value $V_{b,M_{\text{DC}}}(o)$ of a macro-operator o is the expected utility gain for deciding directly after executing the macro-operator o over acting immediately, minus its computational cost. Macro-operators are selected according to Equation 4, and we have also added an error model that chooses macro-operators at random (Equation 3).

Results

Model Comparisons We began by evaluating how well each model explains the aggregate data pooled across all participants. We fit each model's free parameters by maximum likelihood estimation. For the models with graded preferences (bounded-optimal, best-first and directed cognition), we considered temperature values τ from 10^{-5} to 10. For models with binary (yes/no) preferences, τ is redundant with ε; thus, we fixed $\tau = 10^{-10}$ for depth-first, breadth-first, and progressive deepening. For the satisficing parameter, we considered all possible positive path values. For the pruning parameter, we considered all possible negative path values. These constraints rule out the degenerate case in which the model suggests taking the termination action in the initial belief state. The bounded-optimal model is unique in providing its own termination and pruning criteria, making these parameters unnecessary.

To account for the differing number of parameters, we computed the Bayesian Information Criterion (Schwarz, 1978). As shown in Table 1, our data provided strong evidence in favor of the optimal model in terms of the complexity-penalized BIC and the raw likelihoods. Modeling process-tracing data at the resolution of individual planning operations is intrinsically difficult. By normalizing and then exponentiating the log-likelihoods, we find that the bounded-optimal strategy has an effective predictive accuracy of 13.4% compared to 12.4% for best-first and 10.3% for the random model. The rather low predictive accuracy each model achieved individually may be partly due to the individual differences documented below. Next, we characterize in which ways people's strategies deviated from and conformed to bounded-optimal planning.

Qualitative predictions Our bounded-optimal planning strategy predicted that people should plan in a manner similar to best-first search. These predictions were mostly confirmed: The optimal strategy always inspected a node on one of the most promising paths and people did so 82.1% of the time ($p < 10^{-15}$). For instance, the optimal strategy always switches the branch if the observed value is below average but stays on the same branch if the observed value is above average, and people do so 86.4% of the time ($p < .0001$) and 59.7% of the time ($p = .00001$), respectively.

Unlike best-first search, the bounded-optimal strategy terminates search when the expected value of information drops below the cost of attaining that information. This predicts that contrary to the standard notion of satisficing, the optimal aspiration level changes dynamically as people acquire more information. For example, in a state of high uncertainty, one should only terminate planning for an option that is exceptionally good. However, as one accumulates information, the probability of finding a better option diminishes. So, at some point, one is better off settling for a mediocre option than wasting energy grasping at straws. As shown in Figure 2a, holding the value of the best path constant, the probability that the bounded-optimal strategy stops planning increases with the number of clicks already made. This pattern is noisily expressed in the human data as well: A mixed effects logistic regression of the termination probability on the number of revealed states and the value of the current best path revealed a significant negative interaction ($\chi^2(1) = 43.319, p < 10^{-10}$). This suggests that, unlike best-first search, people dynamically adapt their aspiration level as predicted by bounded-optimal planning.

However, we also found systematic deviations of human planning from the bounded-optimal strategy: After observing a large positive reward for taking a certain action in the first step in their first click, people more often conformed to best-first search which evaluates the immediate next step (51.5%) than to the optimal strategy which skips ahead to one of its final destinations (28.9% of the time; $\chi^2(1) = 10.1, p < .0015$). Furthermore, if the first click revealed a small positive reward for one of the first actions, participants inspected the reward of the next step only 25.1% of the time; instead they switched to another branch 57.0% of the time. This is inconsistent with both optimal planning and best-first search. While people

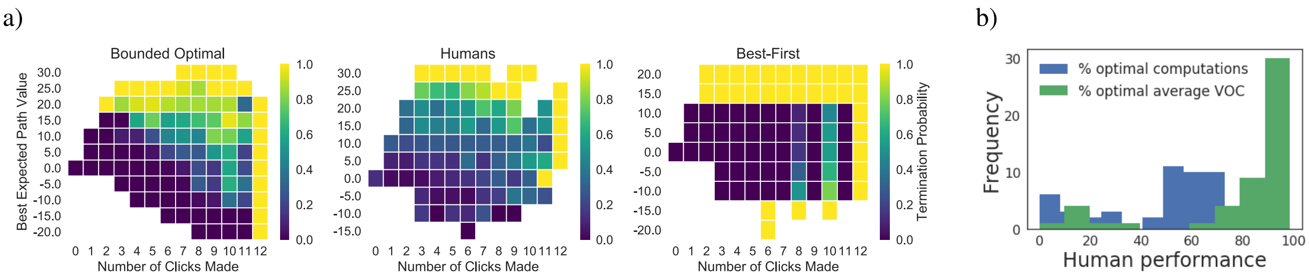

Figure 2: a) Adaptive termination threshold in bounded-optimal and human planning. Both the bounded-optimal model and humans smoothly lower the termination threshold as more clicks are made. The next-best-fitting model does not have this characteristic. b) Individual differences in the rationality of people's planning strategies.

conformed to bounded-optimal planning in that they predominantly started by inspecting the reward of a first step or a second step rather than one of the final rewards (82.6% vs. 17.4%, $p < 10^{-15}$) they preferred inspecting the rewards for the first step over the rewards for the second step (72.6% vs. 10.0%, $p < .0001$) whereas bounded-optimal planning is indifferent between the two. These deviations might reflect that simulating actions in future states is more costly than simulating acting in the present.

Quantifying deviations from bounded optimality. We found that, on average, 45.0% of our participant's computations were sub-optimal. However, the computations people selected did nevertheless achieve 86.8% of the highest possible value of computation ($\text{VOC}(b,c) = Q^{\star}_{m}(b,c) - Q^{\star}_{m}(b,\perp)$). Next, we characterized the ways in which people's planning strategies are sub-optimal. We found that people tend to plan too little. Concretely, 28.3% of people's deviations from bounded-optimal planning are caused by stopping too early, but only 6.3% are caused by stopping too late. Finally, the majority of people's deviations from bounded optimality (i.e., 65.5%) occurred when they clicked on one node when the optimal strategy would have clicked on a different node.

In summary, we found three systematic deviations from bounded optimality: First, unlike the optimal strategy, people preferred to inspect states in the order they would traverse them. Second, people tend to stop planning too early. Third, following a (very) good observation on the first click, people continue to explore broadly when they should zoom in on the most promising paths identified by that observation.

Individual differences in rationality. Consistent with previous work by Stanovich and West (1998) we found considerable inter-individual differences in the extent to which people's planning strategies were rational (see Figure 2b). The average agreement between people's planning operations and bounded-optimal planning ranged from 0% to 80.8% (average 48.0%, standard deviation 23.3%); and the average VOC of people's planning operations ranged from 0% to 98.5% of the optimal VOC (mean 80.0%, standard deviation 26.4%). Figure 2b suggests that the majority of participants used planning strategies that achieve a high level of resource-rationality; the frequency of alternative strategies decreased with the degree to which they were sub-optimal, and the distribution of people's rationality scores might be bimodal.

Experiment 2: Structure shapes strategies

Bounded optimality predicts that people should adapt their planning strategy to the structure of the environment. We test this prediction by manipulating whether future rewards are more variable than immediate rewards or vice versa.

Methods

Experiment 2 presented participants with a modified version of the three-step planning task from Experiment 1 (see Figure 1). Each participant was randomly assigned to one of two conditions that differed in whether the variability of a node's reward distribution either increases or decreases with the number of steps it takes to reach that node. Concretely, in the first condition the reward distributions were $\text{Uniform}(\{-4,-2,+2,+4\})$, $\text{Uniform}(\{-8,-4,+4,+8\})$, and $\text{Uniform}(\{-48,-24,+24,+48\})$ for nodes reachable in one, two, and three steps, respectively. In the second condition, the order of these distributions was reversed. The instructions informed participants about this reward structure. Participants then completed 10 practice trials with fully revealed reward structures in which they could learn the statistics of the environment from experience. Next, participants answered a quiz about the range of rewards at the first step and the third step, the cost of clicking, and their bonus.

We recruited 69 participants on Amazon Mechanical Turk; 16 of them (23%) were excluded for either never clicking during the training block or incorrectly answering more than one of the four quiz questions. Each participant received a base pay of \$0.50 and a performance dependent bonus that was proportional to their final score in the game (avg. bonus: $\$1.84 \pm 0.81$) for about 16.5 minutes of work on average.

Results

Model predictions We computed the bounded-optimal planning strategy for each environment and characterized its behavior. Bounded optimality predicted that people should use qualitatively different planning strategies in these two environments. Broadly speaking, when the variance of the reward distribution increases with the number of steps then people should plan backward from potential end states. In contrast, when the variance decreases with each step, then people should plan forward from their initial state.

Test of qualitative predictions As predicted by our resource-rational analysis, participants engaged in forward planning when the variance of the reward distribution was decreasing and backward planning when it was increasing. Concretely, in the condition with outwardly increasing variance the first click inspected a potential end state 95.0% of the time compared to 0.3% in the condition with decreasing variance ($\chi^2(1) = 520.5, p < .0001$). Conversely, in the decreasing variance condition 99.7% of the first clicks inspected one of the immediate rewards compared to only 4.1% in the increasing variance condition ($\chi^2(1) = 478.3, p < .0001$). Furthermore, when the variance increased outwardly, then only 13.9% of participants inspected any of the rewards at steps 1 or 2 before they had inspected all potential end states. Likewise, when the variance decreased outwardly, 86.1% of participants' second clicks also inspected an immediate reward unless the first click observed the largest possible reward in which case 69.3% of them stopped planning as predicted. Like the optimal strategy, participants in the increasing variance condition stopped 81.6% of the time they discovered a terminal state with the highest possible reward.

Model comparisons We found that our bounded-optimal planning model (BIC = 21203) explained our participants' click sequences in the two conditions substantially better than the directed cognition model (BIC = 25354), the best-first search model (BIC = 29122), and the random model (BIC = 29313). While the bounded-optimal model and the directed cognition model correctly predicted forward versus backward planning, none of the classic models of planning can capture people's backward planning in the increasing variance condition. This highlights the advantage of having a general theory of how people's cognitive strategies are shaped by the structure of the environment and cognitive constraints.

Conclusion

In summary, our resource-rational analysis predicted people's planning strategies more accurately than previous models of planning, and it captured how people's planning strategies depend on the structure of the environment. Furthermore, we have presented the first quantitative assessment of the rationality of people's cognitive strategies against the metric of bounded rationality on a computation-by-computation level.

Our study illustrates the potential of resource-rational analysis for elucidating people's cognitive strategies and understanding why they are used. Our findings suggest that this approach can make valuable contributions to the debate about human rationality by enabling a quantitative assessment of people's cognitive strategies against realistic normative standards and a fine-grained characterization of when and how they deviate from bounded-optimal information processing.

Acknowledgement This work was supported by grant number ONR MURI N00014-13-1-0341 and a grant from the Templeton World Charity Foundation.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Psychology Press.

Callaway, F., Lieder, F., Krueger, P. M., & Griffiths, T. L. (2017). Mouselab-MDP: A new paradigm for tracing how people plan. In *The 3rd multidisciplinary conference on reinforcement learning and decision making.*

De Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Grouton.

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *The American Economic Review*, *96*(4), 1043–1068.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, *7*(2), 217–229.

Hay, N., Russell, S., Tolpin, D., & Shimony, S. (2012). Selecting Computations: Theory and Applications. In N. de Freitas & K. Murphy (Eds.), *Proceedings of the 28th conference on uncertainty in artificial intelligence.* Corvallis: AUAI Press.

Huys, Q. J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, *8*(3), e1002410.

Huys, Q. J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., ... Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, *112*(10), 3098–3103.

Lieder, F., Griffiths, T. L., & Hsu, M. (2017). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review*.

Lieder, F., Griffiths, T. L., Huys, Q. J., & Goodman, N. D. (2017). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, 1–28.

Lieder, F., Krueger, P. M., & Griffiths, T. L. (2017). An automatic method for discovering rational heuristics for risky choice. In *Proceedings of the 39th annual meeting of the cognitive science society. austin: Cognitive science soc.*

Newell, A., & Simon, H. (1956). The logic theory machine–a complex information processing system. *IRE Transactions on information theory*, *2*(3), 61–79.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge University Press.

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129.

Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). Cambridge, MA: MIT press.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of experimental psychology: general*, *127*(2), 161.