AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

2024, Vol. 131, No. 4, 905–951
https://doi.org/10.1037/rev0000456

Identifying Resource-Rational Heuristics for Risky Choice

Paul M. Krueger[1], Frederick Callaway[2], Sayan Gul[3], Thomas L. Griffiths[1, 2], and Falk Lieder[4]

[1] Department of Computer Science, Princeton University
[2] Department of Psychology, Princeton University
[3] Department of Psychology, University of California, Berkeley
[4] Max Planck Institute for Intelligent Systems, Tübingen, Germany

Perfectly rational decision making is almost always out of reach for people because their computational resources are limited. Instead, people may rely on computationally frugal heuristics that usually yield good outcomes. Although previous research has identified many such heuristics, discovering good heuristics and predicting when they will be used remains challenging. Here, we present a theoretical framework that allows us to use methods from machine learning to automatically derive the best heuristic to use in any given situation by considering how to make the best use of limited cognitive resources. To demonstrate the generalizability and accuracy of our method, we compare the heuristics it discovers against those used by people across a wide range of multi-attribute risky choice environments in a behavioral experiment that is an order of magnitude larger than any previous experiments of its type. Our method rediscovered known heuristics, identifying them as rational strategies for specific environments, and discovered novel heuristics that had been previously overlooked. Our results show that people adapt their decision strategies to the structure of the environment and generally make good use of their limited cognitive resources, although their strategy choices do not always fully exploit the structure of the environment.

Keywords: decision making, heuristics, risky choice, bounded rationality, strategy discovery

We make thousands of decisions every day. Collectively, these decisions determine our personal lives and the success of companies and organizations, and they also shape the economy and society as a whole. However, making good decisions is a challenging computational problem for people and artificial intelligences alike (Bossaerts et al., 2019; Bossaerts & Murawski, 2017; Gershman et al., 2015; Kwisthout et al., 2011; Nowozin, 2014; Papadimitriou & Tsitsiklis, 1986). According to classic economic theory, people should choose their actions to maximize the expected utility of the consequences (Morgenstern & Von Neumann, 1953; Savage, 1951), but computing those expected utilities for real-world problems is a substantial task and humans face significant limitations in computational resources and time (Simon, 1972). As a result, most real-world decisions are too complex for people to apply those economic principles correctly. Instead, people have to rely on heuristics to simplify decision making (Gardner, 2019; Gigerenzer & Goldstein, 1999; Gilovich et al., 2002; Kahneman et al., 1982; Maule & Hodgkinson, 2002).

Despite the ubiquity of heuristics (and resulting biases) in decision making, identifying which heuristics people use and when they use them can be a challenge. Psychologists identify heuristics by thinking about the structure of decision environments and

This article was published Online First April 18, 2024.
Paul M. Krueger https://orcid.org/0000-0001-5698-8984
Sayan Gul is deceased.
This work was supported by Multidisciplinary University Research Initiative (MURI) Grant N00014-13-1-0341 from the Office of Naval Research, Grant FA9550-18-1-0077 from the Air Force Office of Scientific Research, and grants from the Templeton World Charity Foundation and NOMIS Foundation to Thomas L. Griffiths. The authors have no conflicts of interest to disclose.

Preliminary versions of the method and experiment were presented at the 39th Annual Meeting of the Cognitive Science Society, the 3rd Multidisciplinary Conference on Reinforcement Learning and Decision Making, and the 14th Biannual Conference of the German Society for Cognitive Science. This material has been substantially revised and expanded for the present article by collecting a much larger data set, refining the model and the methods used to approximate solutions to the model, and adding the clustering analysis. All code and data used to run the experiments and produce the results presented in this article are available at https://github.com/fredcallaway/rational-heuristics-risky-choice/.

Paul M. Krueger and Frederick Callaway contributed equally to this work.

Thomas L. Griffiths and Falk Lieder share joint senior authorship on this article.

Paul M. Krueger played a lead role in visualization and writing–review and editing and an equal role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, and writing–original draft. Frederick Callaway played an equal role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, visualization, writing–original draft, and writing–review and editing. Sayan Gul played a supporting role in conceptualization, formal analysis, investigation, and methodology. Thomas L. Griffiths played a lead role in funding acquisition and supervision, a supporting role in investigation, methodology, writing–original draft, and writing–review and editing, and an equal role in conceptualization, project administration, and resources. Falk Lieder played a lead role in conceptualization, a supporting role in data curation, formal analysis, supervision, and writing–review and editing, and an equal role in investigation, methodology, project administration, resources, and writing–original draft.

Correspondence concerning this article should be addressed to Paul M. Krueger, Department of Computer Science, Princeton University, 35 Olden Street, Princeton, NJ 08540, United States. Email: paul.m.krueger@gmail.com

observing human behavior, but this process of discovery is slow and requires both luck and ingenuity. This makes discovering good heuristics a critical bottleneck to understanding and improving human decision making. Furthermore, while many specific heuristics have been identified, there is no general method that could be used to predict which heuristics will be used in novel situations.

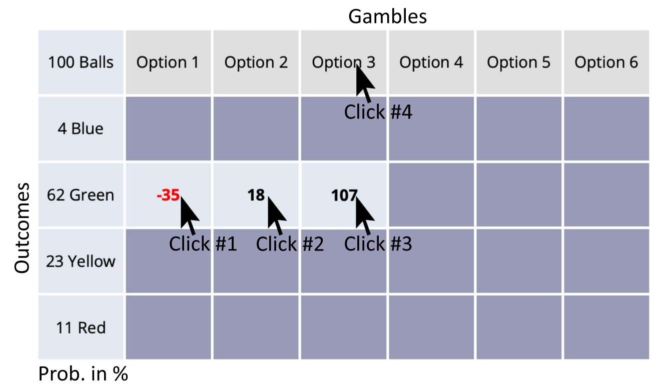
In this article, we address these problems by proposing a theoretical framework that can be used to automatically derive optimal heuristics. This approach relies on the idea that people's heuristics may arise as a rational adaptation to the structure of the environment and the cognitive constraints of limited time and computational resources (Frank, 2013; Griffiths et al., 2012, 2015; Lewis et al., 2014; Lieder & Griffiths, 2020; Simon, 1956, 1972; Zednik & Jäkel, 2016)—a normative benchmark that we refer to as "resource-rationality" (Griffiths et al., 2015; Lieder & Griffiths, 2020). Resource-rationality is achieved through an optimal trade-off between decision quality and computational cost. This trade-off also arises in machines and can be formalized using ideas from the artificial intelligence literature (Russell & Wefald, 1991b). Specifically, heuristic decision making can itself be understood as a sequential decision problem (Griffiths et al., 2019). At each step, people make a decision about whether to collect more information about their options through deliberation or simply to stop thinking and act. Whereas classic rationality applies to the utility of decisions in the external world, and research on heuristics and biases highlights internal cognitive limitations, the framework we propose here bridges these two approaches by viewing rationality as a property of this internal sequential decision process, rather than of the resulting external decisions. We leverage recent advances in machine learning to solve this sequential decision problem, allowing us to automatically derive optimal heuristics for any decision environment.

To demonstrate the accuracy and generalizability of our approach, we applied it to multi-alternative, multi-attribute decision making (Zanakis et al., 1998). The heuristics people use to make these kinds of decisions have been extensively studied in the Mouselab paradigm for multi-alternative risky choice where participants choose between multiple gambles whose payoffs depend on a random outcome (Payne et al., 1988, Figure 1). Participants are shown the probability (prob.) of each outcome and a payoff matrix with one column for each gamble and one row for each outcome. The entry in column *g* and row *o* indicates how much money gamble *g* pays if the 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 strategies. To operationalize the cognitive cost associated with evaluating possible outcomes, participants are charged a fixed fee for every click; thus, to maximize earnings, the player must employ a decision strategy that achieves an optimal trade-off between the cost of information gathering versus the value of information (VOI). Importantly, the Mouselab paradigm thereby allows us to externalize cognitive costs of computation as clicks.

We applied our heuristic-discovery method across a large range of multi-attribute decision-making problems and tested its predictions in an experiment that is an order of magnitude larger than the largest previous study in this setting. Our method automatically rediscovered the classic take-the-best (TTB; Gigerenzer & Goldstein, 1999) heuristic and an information search strategy similar to the weighted-

Figure 1

Illustration of the Mouselab Paradigm (Payne et al., 1988)



Note. The task is to choose one of six gambles, each of which results in one of four probabilistic outcomes; before gambling, participants can gather information about the value of each cell by clicking on it. The Mouselab paradigm externalizes computations by clicks, belief states by revealed information, and the cost of each computation by the fee charged for the corresponding click. This example shows a sequence of clicks generated by the satisficing–take-the-best strategy, which was discovered through our approach. prob. = probability. See the online article for the color version of this figure.

additive (WADD) heuristic (Dawes & Corrigan, 1974; Payne et al., 1988) as resource-rational strategies in specific situations, validating the approach. In addition, our method discovered novel strategies that had been previously overlooked. We collected data from over 2,300 participants, systematically varying the parameters of the decision-making environment. This allowed us to parametrically evaluate human heuristics using the normative standard of resource-rationality. If human heuristics are selected in accordance with this normative standard, people should adapt their strategies to the decision environment.

Our approach correctly predicted which strategies people use and under which environmental conditions they use them more versus less often. Comparing people's strategy choices against the normative standard of resource-rationality indicated that people use resource-rational decision-making strategies and adaptively select which strategy to use based on the structure of the environment. However, they select and execute these strategies imperfectly, thus falling short of perfect resource-rational decision making. In a follow-up experiment, we found that people continued to deviate from resource-rational decision making even when the task was modified such that the assumptions of the resource-rational model were met. Our findings suggest that our automatic strategy discovery method is a promising approach for uncovering people's cognitive strategies and assessing human rationality using a more realistic normative standard.

Background

Before we introduce our approach, we briefly summarize previous work on identifying the heuristics that people use in multi-alternative risky choice and the normative frameworks that have been used to account for these choices.

Manually Identified Heuristics

Previous work has manually identified a number of heuristics employed in multi-attribute risky choice (Gigerenzer & Goldstein, 1996; Katsikopoulos, 2011; Payne, 1976a; Simon, 1956; Thorngate, 1980). Early research focused on additive models in which linear combinations of payoffs are used to make a decision (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975). For example, classical expected utility theory is implemented by the WADD model, in which payoffs are weighted by their probabilities.[1] Another widely recognized heuristic is the lexicographic rule (Svenson, 1979; Tversky, 1969) or "TTB" (Gigerenzer & Goldstein, 1999), which focuses on a single diagnostic attribute. Satisficing, on the other hand, focuses on one alternative at a time, selecting it only if all of its attributes are above a certain cutoff value (Simon, 1956).

Like researchers before them, Payne et al. (1988) studied the trade-off between cognitive effort and decision accuracy afforded by heuristics. They operationalized effort by decomposing heuristics into units of "elementary information processes" (EIPs; Johnson & Payne, 1985). These basic steps of cognitive processing include operations like "read," "compare," "add," "product," "move," and "choose," and this framework has its origins in the view of human reasoners as symbolic information processing systems (Newell & Simon, 1972). Assuming every operation requires equal effort, Payne et al. (1988) reported simulation results showing the effort–accuracy trade-off of nine different heuristics in the Mouselab task. These heuristics included the three aforementioned, two others ("elimination by aspects"; Tversky, 1972 and "majority confirming dimensions"; Russo & Dosher, 1983), and four hybrids or modified versions of the previous five. They showed that certain heuristics require substantially less effort but that, depending on the environment of the Mouselab task, may incur only a minimal reduction in accuracy. For example, when one attribute is much more likely than all the others, TTB performs nearly as well as the much more costly WADD strategy.

Payne et al. (1988) noted general characteristics in the patterns of information processing associated with heuristics. These include the amount of information gathered and the variance in gathering information across attributes versus across alternatives. Rather than measure heuristics directly, they measured these behavioral features in human participants, which we discuss in detail later. They found that people adjust their information processing to the environment, such that less effortful patterns are used when the reduction in accuracy is relatively small and when under time constraints (since less effortful heuristics are simpler and faster).

While appreciating the effort–accuracy trade-off, Payne et al. (1988) assumed that expected value maximization is the normative standard and that heuristics arise as a necessary but suboptimal adaptation to environmental variables. That is, certain heuristics are *less bad* in some environments, but a limitation is all the same. Their simulation results cannot predict which heuristic ought to be used in which environment because EIPs do not specify *how much* effort each operation costs. Rather, the subjective cost of even a single EIP is ultimately a suboptimal cognitive bias. In our work, rather than assume EIPs that have a priori unquantifiable cost, we impose a cost of gathering information directly. This allows us to compute precisely the optimal trade-off between effort (operationalized as click costs) and decision accuracy. In doing so, we provide a normative account of heuristics based on the rational use of costly cognitive operations. In this framework, heuristics can be derived automatically by optimizing the cost–accuracy trade-off, rather than relying on subjective insight to propose or search for strategies.

Normative Accounts of Heuristics

Like Payne et al. (1988), other previous work has also characterized the environments in which hand-crafted heuristics perform best, showing that people select among these heuristics accordingly (Baucells et al., 2008; Dieckmann & Rieskamp, 2007; Gigerenzer & Brighton, 2009; Goldstein & Gigerenzer, 2002; Katsikopoulos, 2011; Katsikopoulos & Martignon, 2006; Martignon & Hoffrage, 1999, 2002; Şimşek, 2013). While challenging classic rationality, this work generally views heuristics as adaptive to the environment rather than adaptive to inherent constraints on the decision-making process itself.

Researchers have previously considered the ideal observer perspective for rational decision-makers (Fishburn, 1989; Geisler, 1989; Howard, 1968), but such an approach was recognized as infeasible (Bell et al., 1988; Kimball, 1958; Simon, 1990; Tversky & Kahneman, 1974). An alternative view is to emphasize the limitations of the decision-maker and the fact that heuristics are computationally cheaper (Payne et al., 1988, 1993) and may achieve some trade-off between accuracy and effort (Beach & Mitchell, 1978; Shah & Oppenheimer, 2008) or optimization under constraints due to information costs (Anderson, 1991; Stigler, 1961), although these perspectives typically view heuristics as inferior to rational decisions (Keeney et al., 1993; Tversky, 1972). The discovery that simpler regression models may outperform more complex ones (Dawes, 1979; Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Schmidt, 1971), combined with observations that heuristics often work quite well in many real-world decision environments (Chater et al., 2003; Czerlinski et al., 1999; DeMiguel et al., 2009; Gigerenzer, 2008; Lee et al., 2002; Lichtenberg & Şimşek, 2017; Wübben & Wangenheim, 2008)—the so-called "less-is-more" effect—challenged the classical normative view of rationality. This led to the idea of ecological rationality (Gigerenzer & Gaissmaier, 2011; Gigerenzer & Todd, 1999; Payne et al., 1993) and attempts to formally account for the effectiveness of heuristics in terms of a range of factors: the structure of the decision environment (Baucells et al., 2008; Bhatia & Stewart, 2018; Dieckmann & Rieskamp, 2007; Katsikopoulos, 2011; Katsikopoulos & Martignon, 2006; Martignon & Hoffrage, 1999, 2002; Şimşek, 2013), trading-off utility and search costs (Analytis et al., 2014) or accuracy and time (Hawkins & Heathcote, 2021; Jarvstad et al., 2012; Rae et al., 2014), bounded evidence accumulation (Brown et al., 2009; Lee & Cummins, 2004), reducing model parameters to balance the bias–variance trade-off (Gigerenzer & Brighton, 2009; Holte, 1993) or limited data (Hogarth & Karelaia, 2005, 2006, 2007; Şimşek & Buckmann, 2015), and using strong priors (Parpart et al., 2018). More recently, a resource-rational analysis of cognition has been applied to view heuristics as making

[1] The traditional notion of expected value maximization under risky choice can be traced all the way to the foundations of probability theory (Huygens, 1657, 1714), while the idea that people instead use subjective utility began with Bernoulli (1738, 1954). Modern research using weighted additive models of utility applied to risky decision making began with Von Neumann and Morgenstern (1944), while Payne et al. (1988) were the first to apply this benchmark to the Mouselab task. For a discussion of the origins of early research on models of risky decision making, see Edwards (1954).

rational use of limited computational resources (Bhui et al., 2021; Binz et al., 2022; Lieder & Griffiths, 2017; Lieder & Griffiths, 2020).

Our approach extends these previous results by automatically discovering the best-performing heuristics by explicitly optimizing over an immense, combinatorial strategy space defined by a set of basic cognitive operations, reminiscent of EIPs (Johnson & Payne, 1985). Expressing heuristics as a rational trade-off between expected payoff and cognitive cost makes it possible to use methods from machine learning to find a near-optimal policy for selecting which costly cognitive operation to perform next given the result of previous operations. In addition to uncovering new heuristics, this approach can establish a normative basis for heuristics that people are already known to use. Any heuristic that our method identifies is likely to strike a near-optimal trade-off between cognitive cost and decision quality.

Automatically Deriving Resource-Rational Heuristics

Our approach rests on the key insight that the process of making a decision can itself be described as a sequential decision problem. At each step of this problem, the agent chooses whether to perform some computation or to instead take the results of previous computations and act. Stated in these terms, the problem of making a decision can be recognized as a Markov decision process (MDP; see Figure 2). A decision-making strategy (a heuristic) is then a policy for that MDP, that is, a function that selects which computation to execute next given the results of previous computations. In the artificial intelligence literature, this problem of choosing a sequence of computations to perform has been formalized as a "meta-level" MDP (Hay et al., 2012) where the name acknowledges that we are deciding how to decide.

The definition of a meta-level MDP parallels that of a conventional, or "object-level" (Russell & Wefald, 1991a), MDP. In an object-level MDP, the environment is represented using states that the agent can occupy and actions that the agent can execute, which lead to rewards and transitions to new states. The agent's objective is to select actions that maximize cumulative reward (Sutton & Barto, 2018). The reinforcement learning literature relies on the MDP framework as a formal representation of the external environment and a source of hypotheses about how to solve the challenges it poses. This has led to considerable recent advances in artificial intelligence (e.g., Berner et al., 2019; Hessel et al., 2018; Mnih et al., 2015; Silver et al., 2017) and success in describing human (e.g., Cohen & Ranganath, 2007; Shteingart & Loewenstein, 2014) and animal (e.g., Rescorla, 1972; Sutton & Barto, 1990) behavior and brain function (e.g., Botvinick et al., 2009; Dayan & Daw, 2008; Glimcher, 2011; Ludvig et al., 2011; Niv, 2009; Schultz et al., 1997).

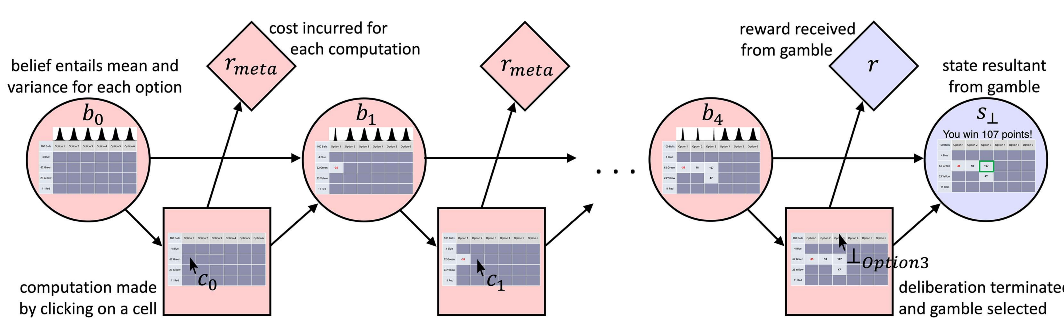
A meta-level MDP uses the same formal framework, but instead of capturing the *external* environment in which decisions take place, it represents the *internal* environment of the cognitive processes that underlie those decisions. As shown in Figure 2, internal states are referred to as *beliefs*, b, and internal actions are described as *computations*, c, that can be used to update beliefs. Because brains and machines have limited computational resources, computations come with a cost, r_{meta}. In addition to making internal computations, an agent can execute a special internal action, $\perp$, that terminates deliberation and takes the action in the external world with the highest expected value according to their current beliefs. The agent then receives a reward from the external world (blue nodes in Figure 2). To identify the best policy for the meta-level MDP, we use methods from reinforcement learning that are used to solve MDPs. This provides a normative account of how a decision-maker ought to navigate the internal world of their mind. In this way, a meta-level MDP can be used to derive cognitive strategies for decision making.

The meta-level MDP has its origins in the artificial intelligence literature on rational metareasoning (Hay et al., 2012; Russell & Wefald, 1991b), which is concerned with building machines that best use their limited computational resources. Recently, however, the approach has been applied to understand how humans efficiently use their cognitive resources. In particular, meta-level MDPs have been used to build resource-rational models of simple (non-multi-attribute) decision making (Callaway, Rangel, & Griffiths, 2021) as well as planning (Callaway, Lieder, et al., 2022; Callaway, van Opheusden, et al., 2021). Here, we apply this approach to compute resource-rational heuristics for multi-attribute risky choice and compare them to the strategies that people use.

Our approach builds on previous work modeling heuristics in decision making in terms of EIPs (detailed above; Bettman et al.,

Figure 2

Schematic Illustration of the Meta-Level Markov Decision Process Framework Applied to the Mouselab Task



Note. At the beginning of each trial, when all cell values are hidden, the agent's initial belief state, b_0, is represented as Gaussian distribution for each of the six gambles. Each time the agent makes a computation, c, by clicking on a cell to gather information, it incurs a computational cost, r_{meta}, and updates its belief distribution for the observed column. When the agent is finished gathering information, it can choose to terminate deliberation, $\perp$, by selecting a gamble, at which point an action is taken in the external world and it receives a reward (blue nodes). See the online article for the color version of this figure.

1990; Johnson & Payne, 1985; Payne et al., 1988). Like this previous work, we model the decision-making process as a sequence of simpler cognitive operations. However, unlike previous work, we do not manually specify how the operations should be sequenced; instead, we derive optimal sequences automatically. That is, we pose the sequencing problem as a meta-level MDP and identify a near-optimal policy that chooses which operation to perform next given the outcome of previous operations. This allows us to exhaustively explore the space of heuristics, identifying those that are adaptive in specific circumstances, rather than relying on human creativity to generate hypotheses about the heuristics people might follow.

Solving complex meta-level MDPs is a challenging computational problem whose complexity exceeds the capacities of standard methods from reinforcement learning and dynamic programming. To overcome this challenge, we recently developed a new reinforcement learning algorithm that is specifically tailored to solving meta-level MDPs called Bayesian meta-level policy search (BMPS; Callaway et al., 2018). Here, we use this technical advance to discover rational heuristics for risky choice. The resulting approach is as follows: First, we model the distribution of decision problems posed by the environment and the cognitive capacities the decision-maker has available to solve those problems as a meta-level MDP. Next, we apply BMPS to solve the meta-level MDP. Last, we characterize this solution in terms of discrete decision strategies by applying a clustering algorithm to the cognitive operations it performs to make its decisions.

Automatically Discovering Strategies for Mouselab

We set out to discover resource-rational heuristics by applying our automatic strategy discovery method to the Mouselab task, the classic process-tracing paradigm for multi-attribute risky choice described above. In the experimental task (illustrated in Figure 1), participants must select from a set of six gambles with four possible outcomes. To reveal the value of a given gamble under a given outcome, participants must click the corresponding cell in a table, paying a cost for doing so. As illustrated in Figure 2, we model this task as a meta-level MDP in which the belief state captures a distribution for the expected value of each gamble given the currently revealed values and computations correspond to revealing a cell and updating the expected value distribution accordingly. Solving this meta-level MDP yields a decision-making policy that optimally trades off between the costs and benefits of considering additional information.

The following sections explain how we modeled the problem of meta-decision-making in the Mouselab paradigm as a meta-level MDP, how we solved this problem to identify optimal strategies, and how we characterized the resulting solutions in terms of simple heuristics.

The Mouselab Paradigm

In our version of the Mouselab paradigm, the alternatives are gambles and the attributes of each gamble are its payoffs in the event of different outcomes. The Mouselab paradigm traces people's decision process by recording the order in which they inspect different pieces of information. Concretely, participants are presented with a payoff matrix where the columns correspond to the alternatives they are choosing between and the rows correspond to different outcomes. Each cell in the payoff matrix specifies how much the alternative corresponding to its column would pay (in points, which translate to a monetary payoff) if the event corresponding to its row was to occur. Critically, all the payoffs are initially occluded, and the participant has to click on a cell to reveal its entry. Each click comes at a cost, and participants are free to inspect as many or as few cells as they would like. The probabilities of the different outcomes are displayed from the beginning of the trial.

Importantly, we view the task as an externalization of a decision-making process that would typically occur mostly or entirely in a person's mind. Thus, clicking a cell in the matrix corresponds to the cognitive operation of evaluating a possible outcome (e.g., by memory look up or simulation). Because externalizing this operation removes most of the associated cognitive cost, we externalize that cost as an explicit point cost.

The resource-rational model makes strong predictions about how the structure of the environment affects the heuristics people should use. To test these predictions in a systematic and comprehensive way, we considered a wide variety of decision environments that varied across three parameters: (a) the "stakes" of the decision (the variance of possible payoffs), (b) the "dispersion" of the outcome distribution (lower values resulting in more similar probabilities for each outcome), and (c) the "cost" of computation (the number of points subtracted for each click). We considered two levels of stakes and five levels for dispersion and cost, resulting in a total of 50 conditions. Each environment was generated by sampling from a distribution specified by the corresponding condition. The two levels of stakes determined the distribution of payoffs, with lower variation in points for low stakes and higher variation in points for high stakes (points drawn from $\mathcal{N}(0, \sigma^2)$ where $\sigma \in \{75, 150\}$). The five levels of dispersion determined the outcome probabilities, with all outcomes being roughly equally likely for low dispersion, and one outcome being much more likely than others for high dispersion (outcome probabilities drawn from Dirichlet $(\alpha \cdot \mathbf{1})$ where $\alpha \in \{10^{-1.0}, 10^{-0.5}, 10^{0.0}, 10^{0.5}, 10^{1.0}\}$). The cost of collecting information was defined by the number of points subtracted for each click ($\lambda \in \{0, 1, 2, 4, 8\}$).

Meta-Level MDP Model

Before defining our meta-level MDP model, we briefly review generic MDPs (Puterman, 2014). MDPs are the standard formalism for modeling sequential decision problems, in which an agent iteratively interacts with an environment to attain the largest possible sum of rewards. An (undiscounted) MDP is defined by a four-tuple, $M = (\mathcal{S}, \mathcal{A}, T, r)$, where $\mathcal{S}$ is a set of possible environment states, $\mathcal{A}$ is a set of actions that an agent can take, T is a transition function that gives the probability of moving from state $s \in \mathcal{S}$ to state s' conditioned on taking action $a \in A$: $T(s, a, s')$, and r is a reward function describing the reward received for such a transition: $r(s, a)$. A reinforcement learning agent's objective is to learn a policy, π, that maps states onto actions to maximize total expected reward.

A meta-level MDP is a special case of an MDP that is used to describe the sequential decision problem associated with making a decision, through a process of performing computations that update the agent's beliefs about the external world. A meta-level MDP is defined by a four-tuple, $M_{\text{meta}} = (\mathcal{B}, \mathcal{C}, T_{\text{meta}}, r_{\text{meta}})$. Here, states are replaced by a set of beliefs, $\mathcal{B}$, describing what the agent may think; actions are replaced by a set of computations, $\mathcal{C}$, describing cognitive operations the agent can perform; the meta-level transition function, T_{meta}, specifies the probability that a computation, c, made with belief b will lead to a new belief, b': $T_{\text{meta}}(b, c, b')$; finally, r_{meta} encodes both the costs of computation (assigning a negative reward for every computation executed) and also the quality of the ultimate decision (assigning the expected external reward attained for the external action that is ultimately executed; see $r_{\text{meta}}(b, \perp)$ below).

In addition to making computations, at any time, t, the meta-level agent can choose to terminate deliberation by taking action $\perp$, at which point the meta-level reward function, r_{meta}, describes the reward the agent will receive for taking the object-level (i.e., external) action that has highest expected utility given the current belief; thus, $r_{\text{meta}}(b, \perp) = \max_a \mathbb{E}_{s\sim b}[U(s,a)]$ where U is the external utility function. The meta-level agent's objective is to learn a meta-level policy, π_{meta}, that maximizes the trade-off between decision quality, $r_{\text{meta}}(b, \perp)$, and accumulated computation costs, $t \cdot \lambda$, where t is the number of computations executed before termination and λ is the cost of each computation.

We model optimal heuristics for risky choice in the Mouselab paradigm as solutions to the meta-level MDP $M_{\text{Mouselab}} = (B, C, T_{\text{meta}}, r_{\text{meta}})$. Concretely, we characterize the decision-maker's belief state at time t by a set indicating which payoffs have already been observed and processed ($\mathcal{O}_t$) and probability distributions ($b_{t,1}, \ldots, b_{t,n}$) over the expected utilities of the available gambles, each of which is defined by

$$\mathbb{E}[U(g)] = \sum_o p(o) v_{o,g}, \tag{1}$$

where $v_{o,g}$ is the payoff of the gamble g under outcome o (V is the payoff matrix). For each payoff, there is one computation $c_{o,g}$ that inspects the payoff $v_{o,g}$ and updates the agent's belief about the expected value of the inspected gamble accordingly. Since the entries of the payoff matrix are drawn from the Gaussian distribution $\mathcal{N}(\bar{v}, \sigma_v^2)$, the resulting expected value distributions are also Gaussian. Hence, the decision-maker's belief about the expected payoff of the gth gamble is represented by

$$b_{t,g} = \left(b_{t,g}^{(\mu)}, b_{t,g}^{(\sigma^2)}\right), \tag{2}$$

where $b_{t,g}^{(\mu)}$ and $b_{t,g}^{(\sigma^2)}$ are the mean and the variance of the probability distribution on the expected value of gamble g given the belief state b_t. Given the set $\mathcal{O}_t = \{(o^{(1)}, g^{(1)}), \cdots, (o^{(t)}, g^{(t)})\}$ of the indices of the t observations made so far, the means and variances characterizing the decision-maker's beliefs are given by

$$b_{t,g}^{(\mu)} = \sum_o p(o) \cdot \begin{cases} v_{o,g} & \text{if } (o,g) \in \mathcal{O} \\ \bar{v} & \text{otherwise} \end{cases}, \tag{3}$$

$$b_{t,g}^{(\sigma^2)} = \sum_o p(o)^2 \cdot \begin{cases} 0 & \text{if } (o,g) \in \mathcal{O} \\ \sigma_v^2 & \text{otherwise} \end{cases}. \tag{4}$$

That is, the belief about each gamble's value is a Gaussian whose mean is the expected value of the gamble (with unobserved payoffs replaced by the average) and whose variance is the probability-weighted sum of the variance induced by each unobserved payoff. Note that, for simplicity, we assume that utility is linear in points; however, the model could easily be extended to account for nonlinear utility functions (e.g., risk aversion).

The meta-level transition function $T_{\text{meta}}(b_t, c_{o,g}, b_{t+1})$ encodes a probability distribution over what the updated means and variances will be after observing a payoff value $v_{o,g}$ sampled from $\mathcal{N}(\bar{v}, \sigma_v^2)$. The meta-level reward for performing the computation $c_{o,g} \in \mathcal{C}$ encodes that acquiring and processing an additional piece of information is costly. We assume that the cost of all such computations is a constant λ. The meta-level reward for terminating deliberation and taking action is $r_{\text{meta}}(b_t, \perp) = \max_g b_{t,g}^{(\mu)}$, since the agent will choose the action with the gamble with the highest expected value.

Using this formalism, we can define a resource-rational heuristic h^* as the optimal policy for a meta-level MDP. The optimal meta-level policy is the one that maximizes the meta-level reward for making a decision in a well-informed belief state minus the cost of attaining it, that is,

$$h^* = \arg\max_{\pi_{\text{meta}}} \mathbb{E}\left[\sum_t r_{\text{meta}}(b_t, \pi_{\text{meta}}(b_t))\right], \tag{5}$$

$$= \arg\max_{\pi_{\text{meta}}} \mathbb{E}\left[\max_g b_{t_\perp}^{(\mu)}(g) - t_\perp \cdot \lambda\right], \tag{6}$$

where the random variable $t_\perp$ is the time step in which the meta-level policy terminates deliberation and λ is the cost of a single computation. Having redefined resource-rational heuristics in this way now allows us to discover them by solving meta-level MDPs. To be able to solve complex meta-level MDPs, we recently developed the BMPS algorithm (Callaway et al., 2018). In Appendix A we provide details of how this algorithm can be applied to find near-optimal strategies in this model.

Identification of Resource-Rational Strategies

Before we discuss the strategies our method discovered, it is important to note that the Mouselab task only externalizes part of the decision-making process, namely information search. We cannot see how people actually *use* that information. How the acquired information is used differs substantially across previously proposed heuristics. For example, the equal weighting (Glöckner & Pachur, 2012) and WADD strategies both consider all available feature values, but they will often yield different decisions because they integrate the acquired information differently. Given that we cannot directly observe how people use the information they acquire, we chose to focus on the search process itself (assuming, for identifying an optimal search process, that the collected information will be used optimally). Accordingly, when comparing the model predictions with previously proposed heuristics, we abstract away from how information is used and focus instead on which information is used at all.

As discussed above, previous work has identified a set of well-known heuristics that people use in multi-attribute risky choice, each of which is associated with a different pattern of information seeking. For example, TTB chooses between alternative options based on the one single attribute that is the best predictor of the outcome (Gigerenzer & Goldstein, 1996).[2] Another heuristic, satisficing (SAT), considers alternative options until it finds one that

[2] If there is a tie, then TTB considers the second most predictive attribute (and so forth) but this scenario virtually never occurs in our paradigm because there are about 1,000 possible payoffs.

is good enough (Simon, 1956); it is sometimes referred to as a conjunctive rule. These heuristics both ignore information about some alternatives or attributes. In contrast, a less frugal strategy, WADD, considers all the available information and computes the expected payoffs of all alternatives (Gigerenzer & Goldstein, 1999; Payne et al., 1988; Simon, 1956). It remains unknown, however, whether additional heuristics exist. Here we set out to discover new heuristics by exploring the full space of potential information search strategies encompassed by all the wide range of decision environments we considered.

We found the best strategy for each of the 1,000 distinct Mouselab problems, corresponding to 20 random samples of payoff matrices in each of the 50 conditions outlined above. To explore this space in a data-driven way, we applied the *k*-means clustering algorithm to the sequences of actions ("clicks") performed by our resource-rational model. *k*-means clustering partitions the click sequences into *k* discrete clusters of similar sequences, with the centroid of each cluster showing the prototype click sequence for that cluster. These prototypes highlight distinct types of heuristics that may be deployed in the Mouselab task.

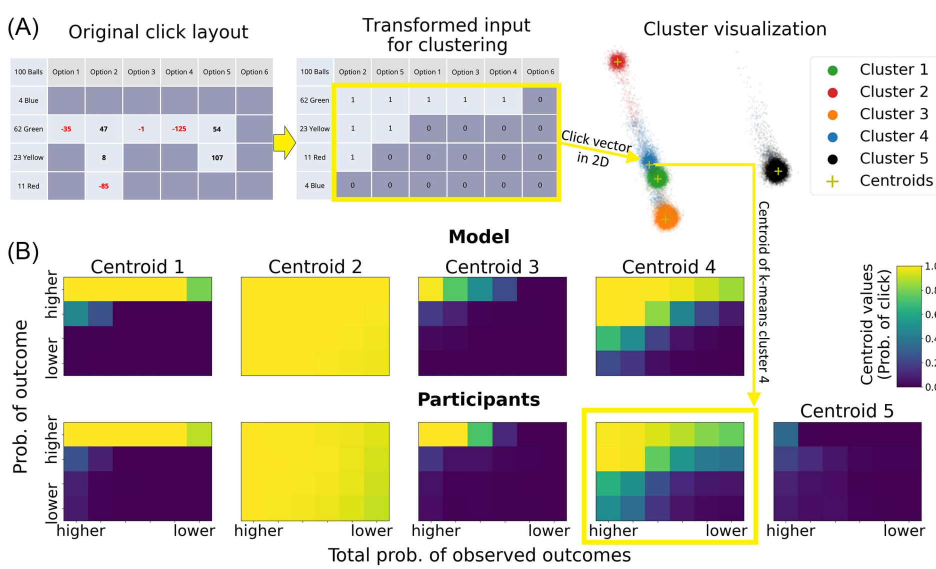
Prior to applying clustering, we transformed the click sequences into a standard format as shown in Figure 3A. The following steps were performed to reduce uninformative spatial variation across trials in the locations of clicks. First, for each problem, a 4×6 indicator matrix of click locations in the Mouselab grid was generated. Second, for each column, the sum of outcome probabilities for every observed cell was computed. Last, we performed the following transformation on the indicator matrix: Rows (outcomes) were rearranged from the most to the least probable outcome, and columns (gambles) were rearranged in descending order of the sum of the probabilities of the outcomes observed in that column. This transformed binary matrix from each trial was collapsed into a vector of length 24 (representing click locations but not the temporal sequence of clicks), which comprised a sample for *k*-means clustering.

We applied the Elkan *k*-means clustering algorithm to the locations of clicks predicted by our resource-rational model across all 1,000 problems, with a Euclidean distance metric (Elkan, 2003). In this and all subsequent analyses, the distribution of the 1,000 problems used to measure the model's behavior was exactly proportional to the particular distribution of those trials received by all participants, to remove variance from model-participant comparisons. We selected $k = 4$ clusters because this identified unique types of click patterns; $k > 4$ resulted in redundant clusters (see Figure B2, for a comparison of different values of *k*). Note that this could be due to a limit in the number of strategies people use or a limitation of the clustering method.

Figure 3B (top) shows the centroids identified in the resource-rational click sequences. Inspecting Centroid 1 suggests that our

Figure 3

Identification of Heuristics



Note. (A) The sequence of clicks on a given trial is converted into an indicator matrix with uninformative spatial variation removed. Rows are rearranged from the most to least probable outcome, and columns are rearranged in descending order of the sum of the probabilities of the outcomes observed in that column. This matrix is then flattened into a 24-dimensional vector. All 47,360 such vectors from our behavioral experiment (2,368 participants × 20 trials per participant; visualized here projected onto 2D space via linear discriminant analysis; Fisher, 1936) serve as input to a *k*-means clustering algorithm. A similar analysis was conducted on the optimal heuristics identified by our model for the corresponding scenarios. (B) Centroids for the clusters uncovered in human data and model simulations from Experiment 1. The first two clusters correspond in certain ways to previously identified strategies: take-the-best (TTB) and weighted additive (WADD), respectively. The third and fourth clusters correspond to the newly discovered strategies: satisficing-TTB (SAT-TTB) and targeted search. A fifth cluster corresponding to gambling randomly (without gathering information) was also revealed in the human data. prob. = probability. See the online article for the color version of this figure.

method rediscovered the TTB heuristic (Gigerenzer & Goldstein, 1999), in which all gambles are evaluated based on the single most likely outcome. This indicates that TTB strikes a near-optimal tradeoff between decision quality and cognitive cost, at least in some situations. Rediscovering this classic heuristic provides support for the validity of our approach. In contrast to the frugal TTB, Centroid 2 corresponds to considering (nearly) all the available information, consistent with the WADD strategy. We refer to this pattern of information seeking as "exhaustive search."

Turning next to Centroid 3, we see a pattern of information search not consistent with any previously proposed heuristic, to our knowledge. We interpret this pattern as the search behavior of a previously unknown heuristic. We call that heuristic "SAT-TTB" because it combines elements of TTB and satisficing (see Figure 1). 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, reducing the amount of information considered.

Finally, Centroid 4 corresponds to an extended version of TTB in which additional attributes of some options are considered. Specifically, it starts by inspecting some or all of the payoffs for the most probable outcome (as in SAT-TTB) but then inspects additional payoffs for the second-most probable outcome from one or more of the most promising gambles. Examples of this strategy are shown in the sequence of clicks illustrated in Figures 2 and 3A. This pattern of information seeking is consistent with both the lexicographic semiorder strategy (e.g., Birnbaum & Gutierrez, 2007; Manzini & Mariotti, 2012; Safarzadeh & Rasti-Barzoki, 2018; Tversky, 1969) and elimination-by-aspects (Tversky, 1972). However, unlike these heuristic decision strategies, the rational model integrates information across attributes, selecting the option with the maximal expected value given all revealed information. We refer to this strategy as "targeted search."

In addition to allowing us to identify these four strategies from the optimal click patterns produced by the model, our approach allows us to generate predictions about when a rational agent should choose to employ each strategy. In particular, the 50 different conditions reflecting different combinations of stakes, dispersion, and cost result in significant variation in which strategy the model predicts should be employed. In the remainder of the article, we compare these predictions against human behavior, allowing us to examine whether people appropriately adapt to which strategy they use and how closely they approximate resource-rational performance.

Experiment 1: Evaluating the Model Predictions

To evaluate the model predictions, we conducted a large-scale experiment, collecting choices from human participants in each of the 50 conditions used to generate our model predictions.

Method

Participants

We recruited 2,368 participants on Amazon Mechanical Turk (1,115 females, $M_{\text{age}} = 37.6$ years, $SD = 16.4$ years) and paid them \$0.50 plus a performance-dependent bonus of up to \$10.38 (average bonus \$3.25) for a mean of 10.2 min of work ($SD = 4.1$ min). Informed consent was obtained using a consent form approved by the Institutional Review Board at Princeton University.

Stimuli and Procedure

Following instructions and a comprehension check, participants performed a variation of the Mouselab task (Payne et al., 1988). Each of the 20 trials began with a 4×6 grid of occluded payoffs: six gambles to choose from (columns) and four possible outcomes (rows). The occluded value in each cell specified how much the gamble indicated by its column would pay if the outcome indicated by its row occurred. The outcome probabilities were described by the number of balls of a given color in a bin of 100 balls from which the outcome would be drawn (see Figure 1). For each trial, participants were free to inspect any number of cells before selecting a gamble. Clicking on a cell revealed its payoff, and participants were charged a fixed cost per click, depending on the condition. The value of each inspected cell remained visible onscreen for the duration of the trial. When a gamble was chosen, participants were informed about which outcome had occurred, the resulting payoff of their chosen gamble, and their net earnings (payoff minus click costs).

The experiment used a $2 \times 5 \times 5$ between-subjects factorial design with a total of 50 conditions, corresponding to those used to generate the model predictions above. The parameters in each condition were the same as those used for model simulations. These parameters included (a) the stakes of the decision, with lower variation in points for low stakes and higher variation in points for high stakes (points drawn from $\mathcal{N}(0, \sigma^2)$ where $\sigma \in \{75, 150\}$), (b) the dispersion of outcome probabilities, with one outcome being much more likely than others for low dispersion and all outcomes being roughly equally likely for high dispersion (outcome probabilities drawn from Dirichlet($\alpha \cdot \mathbf{1}$) where $\alpha \in \{10^{-1.0}, 10^{-0.5}, 10^{0.0}, 10^{0.5}, 10^{1.0}\}$), and (c) the cost of collecting information, defined by the number of points subtracted for each click ($\lambda \in \{0, 1, 2, 4, 8\}$).

The instructions explained the task by walking the participant through the demonstration of a trial with step-by-step explanations. These explanations covered the cost of clicking, the way that their payoff was determined, the range of payoffs, how some outcomes were more likely than others, and a description of the performance bonus (\$0.01 for every 5 points). Participants were given three practice trials, and after these instructions, they were given a quiz that assessed their understanding of all critical information conveyed in the instructions. The full experiment, including instructions, can be viewed at https://kcggl-expt1.netlify.app/. If a participant answered one or more questions incorrectly, they were required to reread the instructions and retake the quiz. If they failed the quiz 3 times, they were not allowed to participate in the main task.

Transparency and Openness

Our results did not exclude any participants (except were noted for comparisons), the sample size per experimental condition was selected prior to data analysis, and we report effect sizes. The model simulations were run using Julia (Bezanson et al., 2017), including the BayesianOptimization library. The behavioral analyses were run using Python 3 (Van Rossum & Drake, 2009), including the statsmodels (Seabold & Perktold, 2010), scikit-learn (Pedregosa et al., 2011), and SciPy (Virtanen et al., 2020) libraries, and using R (R Core Team, 2020) and RStudio (RStudio Team, 2019) with the lme4 library (Bates et al., 2015). The study design and analysis were not preregistered. All code and data used to run the experiments and produce the results presented in this article are available at https://github.com/fredcallaway/rational-heuristics-risky-choice/.

Results

We compared the clusters of click sequences produced by our model to those produced by human participants. To further assess the theoretical predictions of our method, we next examined how these strategies depend on the structure of the environment. We looked at how the resource-rational method adapts strategy use to the statistics of the environment and then compared this to how people's strategies depend on the environment. Finally, we tested additional theoretical predictions about the variability of people's choice behavior and quantified how our participants' choice behavior deviated from resource-rational decision making.

Identification of Strategies

As an initial analysis, we repeated the k-means clustering procedure we used to characterize the different strategies employed by our resource-rational model. Data from each trial were transformed in the same way as the model predictions, and the resulting representations were clustered. For human participants, using $k = 5$ clusters produced distinct click patterns, whereas using $k > 5$ clusters resulted in groups of redundant strategies (see Figure B3, for a comparison of different values of k). The results are shown in Figure 3B.

The first four clusters recapitulate those produced by the model, manifesting the classic TTB strategy and WADD-like exhaustive search as well as the newly discovered SAT-TTB strategy and targeted search. While the resource-rational model never gambles randomly, participants do occasionally gamble without gathering any information; this is captured in Centroid 5. Overall, this analysis suggests that people use similar information search strategies as the resource-rational model.

Based on the clustering solution, we defined five distinct search strategies to be considered in subsequent analyses as follows: (a) targeted search was defined as clicking one or more cells from the most probable row, and one or more cells from one or more additional rows, but never more cells from a less probable row than from a more probable row; (b) SAT-TTB was defined as selecting 1–5 cells from the most probable row, and nothing else, with the final clicked cell having the highest payoff; (c) TTB was operationalized as selecting all six cells from the most probable row, and nothing else; (d) exhaustive search was defined as selecting all 24 cells; (e) a random strategy entailed zero clicks. Finally, we considered a miscellaneous category of other strategies, which were those not consistent with any of the previous five definitions.

It is important to note that the identified clusters do not correspond exactly to the search strategies defined above. In particular, targeted search can often be captured by Clusters 1 or 3, in addition to Cluster 4. To quantify these differences, Figure B1 shows a confusion matrix of cluster labels and decision strategies. Overall, the automatic clustering and hand-defined strategy classification agree 55.4% of the time. Importantly, however, much of the confusion occurs between the SAT-TTB and targeted search strategies. For example, a trial that follows SAT-TTB except for one additional click on another feature would likely be clustered with SAT-TTB but would be classified as a targeted search. Collapsing across these two clusters/strategies, the agreement rises to 73.0%. In the remainder of our analyses, we focus on breaking down responses by the hand-defined decision strategies to maximize the interpretability of our results.

Finally, since the k-means clustering analysis was run at the group level, we additionally measured the similarity between the model and participants for patterns of clicking at the level of individual trials. Comparing the click vectors of participants and the model, we found that 66.8% of participants showed a significant ($p < .05$) sensitivity to the structure of individual trials in the manner predicted by the model. See Appendix C for details.

Comparison of Strategies Across Environments

The clustering results indicate that people use the same types of strategies as the resource-rational model. To determine whether people deploy these strategies rationally, we inspected how the frequency with which people use each strategy depends on the structure of the environment. Consistent with our main predictions, we found that participants adapt their strategies to the environment in much the same way as the resource-rational model (see Figure 4).[3]

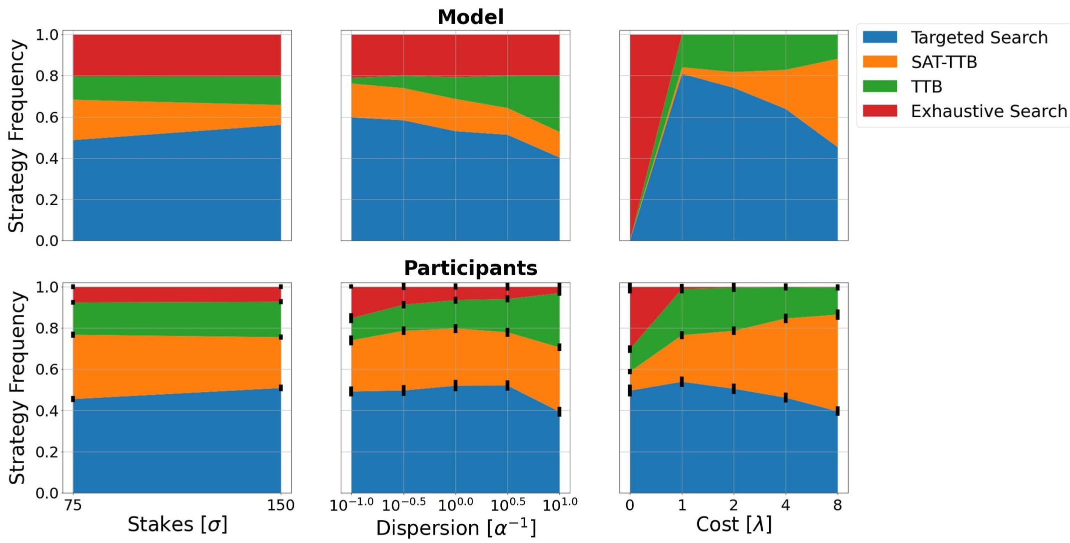
Our resource-rational model predicted that as the stakes increase, participants should rely less on the most frugal strategy—SAT-TTB—and more on targeted search, which gathers additional information. The data confirmed both predictions; that is, regressing the frequencies with which participants used each strategy on each of the three environmental parameters in a logistic mixed-effects regression with random intercepts revealed that the stakes had a significant negative effect on the frequency of SAT-TTB ($B = -2.3$, $p < .001$) and a significant positive effect on the frequency of targeted search ($B = -1.3$, $p < .001$; left panels of Figure 4). In all regressions, B values denote the effect of moving one step up in the condition variable.

The model predicted that as the outcome distribution becomes more peaky (i.e., higher dispersion), the use of TTB should steadily increase; intuitively, one can focus on a single outcome when only one is likely to occur. Our participants confirmed this prediction ($B = -5.5$, $p < .001$; middle column of Figure 4). However, while the resource-rational model most often uses targeted search in low-dispersion environments, participants often resorted to choosing randomly instead ($B = -1.4$, $p < .001$).

When there is no cost for gathering information, the model always uses exhaustive search since the VOI is always positive. Although participants also limited their use of exhaustive search to this case, they were more likely to use targeted search. As the cost increases from 1 to 8, the resource-rational model and participants show the same pattern for the remaining three strategies: decreasing the use of both targeted search ($B = -0.8$, $p < .001$) and TTB ($B = -3.9$, $p < .001$), while increasing the use of the most frugal strategy, SAT-TTB ($B = -3.3$, $p < .001$). Figures D1 and D2 compare strategy frequencies in each of the 50 conditions, showing broad correspondence between the resource-rational model and participants. Table D1

[3] To facilitate the comparison between the model predictions and participant behavior, Figure 4 is conditioned on the four strategies shown, that is, not including undefined patterns of clicking or random gambles.

Figure 4
Strategy Use Across Environments



Note. Use of exhaustive search, take-the-best (TTB), and variations of satisficing-TTB (SAT-TTB and targeted search) by the resource-rational model and human participants in Experiment 1 as a function of the three environment parameters: σ, the standard deviation of possible payoffs, α^{-1}, the peakiness of the outcome distribution, and λ, the cost paid for each piece of information revealed. Error bars show the 95% confidence interval across participants. See the online article for the color version of this figure.

summarizes post hoc pairwise comparisons and effect sizes for the statistics reported in this section.

Understanding Variability in Choice Behavior

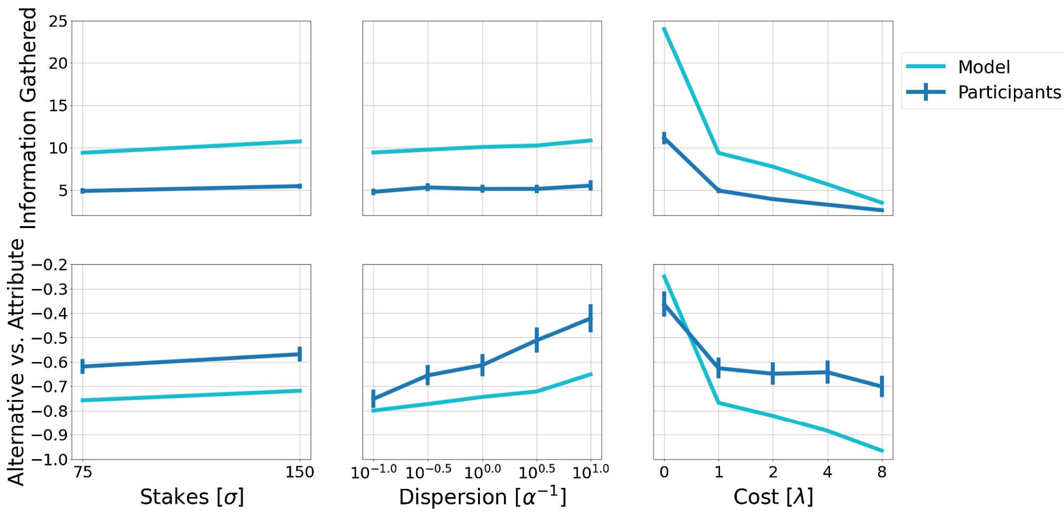
Previous research on multi-attribute risky choice has characterized people's choice behavior in the Mouselab paradigm in terms of four features (Lohse & Johnson, 1996; Payne, 1976b; Payne et al., 1988). The first feature is the total amount of information processed, the second measures the relative frequency of attribute- versus alternative-based information processing, and the third and fourth features measure the variance in information gathering across attributes and alternatives, respectively. Payne et al. (1988) used these measures to assess how participants trade off effort and accuracy across nine hand-selected heuristics, finding that both high dispersion and time pressure lead to less information gathering, more attribute-based processing relative to alternative-based processing, and more selectivity for attributes (i.e., greater variance in information gathering across each). The resource-rational model predicts all of these effects (with click cost having a similar effect as time pressure) as well as a similar pattern when the decision stakes decrease. Here, we confirm that all these effects hold across a broad set of decision environments. However, both the resource-rational model and our participants deviate from the finding of Payne et al. (1988) on the effect of dispersion on alternative variance.

We first considered the total amount of information gathered (i.e., the number of clicks made). As illustrated in Figure 5 (top panels), participants adapted the amount of information gathered to the environmental structure in much the same way as the model, but they consistently gathered too little information. When the stakes increase, the potential for large gains and large losses goes up, and this merits more information gathering. Indeed, participants gathered more information as the stakes increased (a linear mixed-effects regression with random intercepts for participants revealed that the stakes significantly predicted information gathered: $B = 0.57$, $p = .009$). When the dispersion of outcome probabilities increases, people should gather less information, since fewer outcomes (and thus cells) are relevant to each gamble's value; participants trended in this direction ($B = -0.13$, $p = .097$). Finally, people reduced information gathering as it became more costly to do so ($B = -1.9$, $p < .001$). However, across all conditions, participants made on average 4.9 fewer clicks than the resource-rational model. We explore possible explanations for this discrepancy below.

We next looked at a behavioral feature that characterizes the sequences of information gathering. Specifically, we computed a metric that measures the relative frequency of alternative-based versus attribute-based processing. In attribute-based processing, sequential clicks are made on one row/outcome (as in TTB and SAT-TTB); this corresponds to comparing several gambles along one dimension. In alternative-based processing, sequential clicks are made on one column/gamble; this corresponds to evaluating one gamble based on multiple features. We can measure the relative frequency of alternative-based versus attribute-based processing in a given trial as the number of sequential transitions between alternative-based clicks minus the number of sequential transitions between attribute-based clicks, divided by the sum of the two terms (Payne, 1976b; Payne et al., 1988). This yields a number between -1 and $+1$, with positive values indicating alternative-based processing and negative numbers indicating attribute-based processing. Figure 5 (bottom panels) shows that both the model and participants rely more on attribute-based processing overall, but with the model favoring this type of processing more heavily than people. Furthermore,

Figure 5

Behavioral Correspondence Between Participants and the Resource-Rational Model in Experiment 1



Note. Top panels: The average number of values revealed by participants and the model as a function of each environment parameter. Bottom panels: The same, but for a measure of alternative- versus attribute-based processing (negative indicates attribute-based). Error bars show the 95% confidence interval across participants. See the online article for the color version of this figure.

participants adapted their processing pattern to the environment in all of the ways predicted by the model: They used more alternative-based processing as the stakes increased ($B = 0.052$, $p = .016$) and as dispersion increased ($B = 0.084$, $p < .001$), and they used more attribute-based processing as cost increased ($B = -0.073$, $p < .001$). A comparison of information gathering and alternative- versus attribute-based processing for the model and participants across each of the 50 decision environments is shown in Figure E1, showing an overall qualitative correspondence.

Two additional informative behavioral markers are the variance in the amount of information gathered across outcomes and gambles. Attribute variance in information gathering is defined as the variance of the proportion of clicks made on each row/outcome, being zero if clicks are evenly divided across outcomes. High attribute variance is a signature of "noncompensatory" strategies that focus attention on a subset of attributes (because the less important attributes cannot "compensate" for the more important ones; Payne, 1976b; Payne et al., 1988). Alternative variance in information gathering is defined in the same way, but for columns. High alternative variance is a signature of strategies that either gather more information for high-value gambles (as in targeted search) or stop searching once a high-value gamble is found (as in SAT-TTB). Figure E2 shows qualitative correspondence between participants and the resource-rational model for both of these measures. As the stakes increase, both the resource-rational model and the participants spread their clicks more uniformly both across attributes (attribute variance: $B = -0.01$, $p < .001$) and alternatives (alternative variance: $B = -0.004$, $p = .0016$), likely due to an overall increase in information gathering. When one outcome was much more likely than all others, people tended to compare many alternatives on that single outcome without considering any other outcomes. As predicted, increasing the differences between the probabilities of different outcomes (higher dispersion) therefore made people distribute their attention less evenly across the different attributes ($B = 0.0091$, $p < .001$) and more evenly across the alternatives ($B = -0.0026$, $p < .001$). Finally, increasing the cost of information made people more discerning in how much attention they paid to different attributes ($B = 0.017$, $p < .001$) and different alternatives ($B = 0.009$, $p < .001$). Payne et al. (1988) predicted that time pressure would have a similar effect but observed a null result on alternative variance. As noted below, this may be due to a small sample size in their study. Figure E3 shows the qualitative correspondence between the model and participants for these two measures across all 50 decision environments.

It is noteworthy that whereas Payne et al. (1988) observed more selectivity for alternatives (higher alternative variance) with high dispersion, our resource-rational model makes the opposite prediction and this prediction is confirmed by participant behavior. Our prediction makes sense intuitively: As dispersion increases, information from less likely attributes becomes less useful, and therefore multiple samples within a single alternative become less useful, consistent with more frequent use of TTB and less frequent use of targeted search as dispersion increases (middle panels of Figure 4). The most likely explanation for this inconsistency is that the result of Payne et al. (1988) was spurious, as the study included only 16 participants and the p value was between .01 and .05. Table E1 presents the results of additional statistical tests for the results reported in this section.

Decision Quality

In addition to providing a framework for discovering strategies, our formalism provides a realistic normative standard for human decision making. This allows us to determine to which extent human deviations from perfectly rational decision making can be attributed to resource-rational consideration of the cost of gathering information versus genuinely irrational use of one's cognitive resources.

To quantify decision quality on a given trial, we divide the expected value of the chosen gamble by the maximum expected value of any gamble. This relative measure allows us to compare the decisions of participants and the resource-rational model to unboundedly optimal decisions made with perfect information. Note that this metric applies to the decision itself, not the strategy; it thus does not include or account for the cost of information gathering.

As illustrated in Figure 6, our resource-rational model achieved a decision quality of relatively close to the unboundedly optimal standard of 1.0, falling shorter when less information is gathered. Concretely, the average decision quality was 0.895. Furthermore, our model accurately predicted that participants' decision quality increases with the stakes ($B = 0.039$, $p = .0035$), decreases with the dispersion of the outcome probabilities ($B = 0.059$, $p < .001$), and decreases with the cost of gathering and processing information ($B = -0.063$, $p < .001$). These results are shown in Figure 6.

It is apparent that participants underperform compared to the model. The average decision quality of all participants was only 0.542. Thus, 23% of the gap between all participants' decision quality and the decision quality of the unboundedly optimal decision strategy (i.e., maximizing expected value) can be explained by resource-rational sensitivity to the imposed click cost, whereas 77% is due to people's deviations from the resource-rational model. Importantly, this proportion could be further reduced by accounting for additional costs and constraints not considered by our model, which we set out to do in Experiment 2.

Table F1 summarizes the main effects, corrections for multiple comparisons, and effect sizes for measuring decision quality across conditions. Figure F2 shows a qualitative correspondence between participants' and the resource-rational model's decision quality across all 50 environmental conditions; see Appendix F for detailed results when excluding low-effort participants who gambled randomly on more than half of all trials (16.6% of participants).

Net Performance

The decision quality metric plotted in Figure 6 omitted the costs of information gathering. This facilitates comparison across conditions and provides a useful measure of the extrinsic quality of participants' decisions. However, to assess the resource-rationality of our participants' strategies, we need to account for both external rewards (gamble payoffs) and cognitive costs (operationalized here as click costs). We thus define *net performance* on a given trial as the expected payoff of the chosen gamble minus the costs of information gathering (again, normalized by the expected maximum gamble value to account for differences across conditions).

On average, participants' net performance was 61.4% (95% CI [58.8, 61.9]) that of the resource-rational model. That is, our participants were quite far from the resource-rational benchmark. In the following section, we consider possible explanations for this gap.

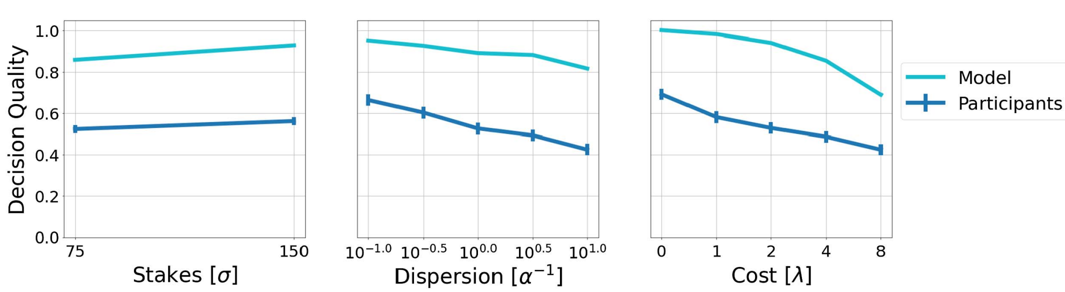
Sources of Underperformance

We identified four possible reasons why people might not meet the performance of the resource-rational model: implicit costs of information gathering, imperfect use of the gathered information, imperfect strategy selection, and imperfect strategy execution. As shown in Figure 7, these four sources, respectively, accounted for 4.2%, 4.2%, 25.0%, and 5.2% of the total performance gap. Note that, as defined, the factors necessarily account for the full gap of 38.6%. See Appendix F for detailed results when excluding low-effort participants.

First, participants may be influenced by costs not accounted for by our resource-rational model, for example, the time required to move the cursor and the anticipated cognitive costs associated with processing the revealed information (Payne et al., 1988). Such a cost could explain why participants collected less information than in the resource-rational model. To assess the degree to which participants' suboptimal performance resulted from insufficient information gathering, we used the performance of a model with an implicit cost parameter fit to match the average number of clicks that participants made (2.4 points per click). The net performance of this model (not including the implicit cost) was 4.2% less than the original model. Thus, implicit costs appear to explain a relatively small portion of the gap.

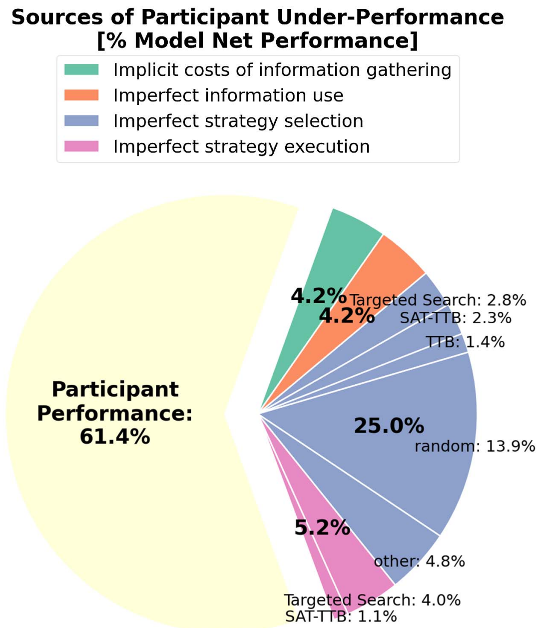
A second source of underperformance is the imperfect use of gathered information. Indeed, participants failed to choose the gamble with the highest subjective expected value on 27.3% of all trials. However, they lost only 10.4 points on average on such trials (i.e., they tended to choose a gamble with close-to-maximal

Figure 6
Decision Quality Across Conditions in Experiment 1



Note. Decision quality is defined as the expected value of the chosen gamble divided by the maximum expected value of any gamble (not including click costs). Error bars show the 95% confidence interval across participants. See the online article for the color version of this figure.

Figure 7
Sources of Underperformance in Experiment 1



Note. Participants' net performance was 61.4% (95% CI [58.8, 61.9]) that of the model, with four distinct sources of the remaining 38.6% gap depicted here. Note that the net performance measure considered here includes both expected payoffs and click costs. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

expected value). In total, imperfect use of information accounted for 4.2% of the gap, about the same portion as implicit costs.

A third possible source of underperformance is imperfect strategy selection. At an aggregate level, people use the same strategies as the model in roughly correct proportion for each environment. However, on a trial-by-trial basis, they may not always choose the most effective strategy. To quantify the impact of imperfect strategy selection on performance, we measured the reduction in net performance on trials in which participants chose a different strategy than the model, while controlling for the previous two sources of underperformance (by conditioning on the amount of information gathered and the use of information).[4] This reduction accounted for 25.0% of the performance gap, by far the largest portion. Notably, random gambling accounted for more than half of this portion of the gap (13.9%).

Finally, even when participants choose the same strategy as the model, they may not execute it perfectly. For example, they may set an incorrect satisficing threshold in SAT-TTB, or they may consider too many or too few additional features in targeted search. To quantify the impact of imperfect execution on performance, we compared the participants' and the model's net performance when there was agreement in trial-wise strategy selection, again controlling for the first two sources of underperformance. This accounted for the remaining 5.2% of the gap. Errors in executing targeted search—the most complicated strategy—accounted for most of this source (4.0%).

Figure 8 provides a more detailed look at the impact of imperfect strategy selection and execution, showing the average reduction in performance associated with each possible model–participant strategy pair. Off-diagonal values correspond to imperfect strategy selection. For example, trials in which participants gamble randomly and the model chooses targeted search account for 7.5% of underperformance, and the sum of off-diagonal values in the "random" column equals the corresponding 13.9% displayed in Figure 7. On-diagonal values correspond to imperfect strategy execution. For example, when both participants and the model chose targeted search, participants lost an average of 4.0% of the model's net performance.

Overall, these results suggest that while people use resource-rational decision strategies and adapt them to the environment in a similar way as the resource-rational model, they often do not use the optimal strategy on a trial-by-trial basis.

Discussion

While our resource-rational model successfully predicted how participants adapt their decision strategies and other behavioral measures to the statistics of the decision environment, they still fell considerably short of the standard set by our model. We have attempted to understand the origins of this underperformance.

The first source of underperformance—implicit costs of gathering information—was measured by controlling for the amount of information gathered by the model. The parameter for the implicit cost of information gathering is meant to account for all additional costs of gathering and processing one piece of information people might experience. This approach assumes that people plan rationally, subject to their cognitive costs. However, it is also possible that people simply gather less information than they should. Furthermore, a simple cost-per-click is only a rough approximation of the true information processing costs (which likely vary depending on which information was acquired). Better characterizing the computational costs involved in risky choice, and dissociating implicit costs from suboptimal information gathering, is an important direction for future research. Finally, the use of clicks to objectify the cognitive cost of information gathering was an important feature of our paradigm, and in the Limitations section, we discuss how future work might refine this procedure.

Experiment 2: Reducing Cognitive Constraints

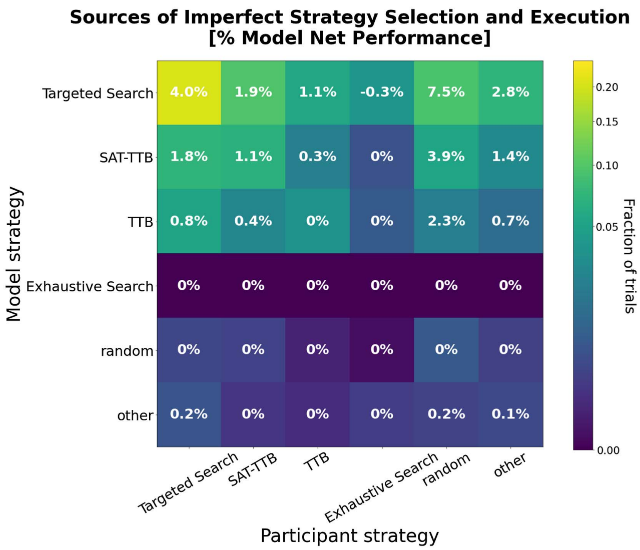
In Experiment 1, we found a substantial gap between the performance of our participants and the resource-rational model. However, it remains unclear to what extent this gap is due to true limitations in human resource-rationality as opposed to a failure of the model to account for all relevant cognitive and opportunity costs. The ideal way to answer this question would be to modify the model to account for these costs; however, there are substantial methodological challenges with doing so (see the Limitations section). Thus, we instead take an empirical approach, modifying the experiment to mitigate some of those costs. If the performance gap is reduced, we can then account for that portion of the gap to the costs that we mitigated.

In particular, we focused on the opportunity cost of time and the cost of computing and maintaining expected values in working

[4] Since the model was simulated 1,000 times per trial, it may occasionally choose different strategies for the same trial. Therefore, the contribution of each strategy to the model's net performance on a given trial—and the extent to which it agrees with participant strategy selection—is weighted by the probability of choosing each strategy on that trial.

Figure 8

Sources of Imperfect Strategy Selection and Execution in Experiment 1



Note. Each cell states participants' average reduction of net performance from a trial-wise comparison of model–participant strategy selection. Off-diagonal cells correspond to imperfect strategy selection, while on-diagonal values correspond to imperfect strategy execution. Colors correspond to the number of trial-wise model–participant strategy pairs. For example, the upper-left cell shows that trials in which participants and the model both selected targeted search contributed to 4.0% to the decrement of participants' net performance (with 9,625 such trials occurring out of the 47,360 trials across all participants, thus the yellow color). The cell just below that shows that participants on average lost 1.8% when they selected targeted search but the model chose SAT-TTB, with 3,394 such trials occurring (thus the teal color). TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

memory. For the former, we required participants to spend a minimum of 20 s on each trial. For the latter, we displayed the subjective expected value of each gamble given the observed information, updating the values whenever new information was revealed.

Method

Participants

We recruited 404 participants on Amazon Mechanical Turk (250 males, $M_{\text{age}} = 37.5$ years, $SD = 10.8$) and paid them \$0.50 plus a performance-dependent bonus of up to \$4.23 (average bonus \$1.66) for about 13.3 min of work on average ($SD = 6.4$ min). Informed consent was obtained using a consent form approved by the Institutional Review Board at Princeton University.

Stimuli and Procedure

The experiment used a $2 \times 2 \times 2$ between-subjects factorial design with a total of eight conditions. The factors we varied between participants were the dispersion of outcome probabilities ($\alpha \in \{10^{-0.5}, 10^{0.5}\}$), the cost of collecting information ($\lambda \in \{1, 4\}$), and whether the participant was in the experimental group or the control group. The stakes of the decisions were low in all conditions ($\sigma = 75$).

For the control group, the task and the instructions were identical to the previous experiment. For the experimental group, the subjective expected value of each gamble given the observed information was displayed next to the label for each gamble. Thus, each time a participant clicked on a cell to reveal its value, the expected value for that gamble was updated according to Equation 3 and displayed atop that column. Furthermore, the experimental group was forced to spend a minimum of 20 s on each trial, and a countdown timer was displayed for the first 20 s of each trial. After the first 20 s, participants were free to spend additional time if they so chose. Figure 9 shows a screenshot from a trial of the experimental condition. These two features of the task were incorporated into the instructions received prior to the task for this group. As a result of these differences, participants in the experimental group spent more time on the task and earned a greater performance bonus on average (16.9 ± 5.3 min, \$1.77 ± \$0.97) than participants in the control group (9.8 ± 5.2 min, \$1.54 ± \$0.95).

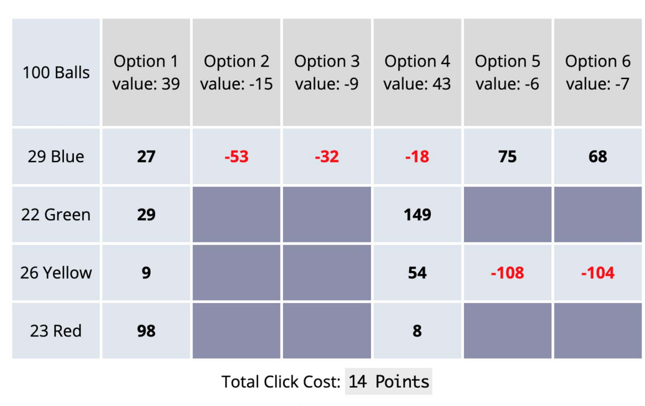
The stakes of the decisions—that is, the variation in outcomes—were always low ($\sigma = 75$). To eliminate variance in performance due to random sampling of trials, we used a single set of 40 problems (20 for each dispersion level), such that every participant in a given condition solved the same set of problems. All participants were required to pass the same comprehension quiz used in the previous experiment. The experimental condition of this experiment can be viewed at https://kcggl-expt2.netlify.app/.

Transparency and Openness

The data analyses relied on all the same practices and software stated for the previous experiment. All code and data used to run the experiments and produce the results are available at https://github.com/fredcallaway/rational-heuristics-risky-choice/.

Figure 9

Screenshot From Experiment 2



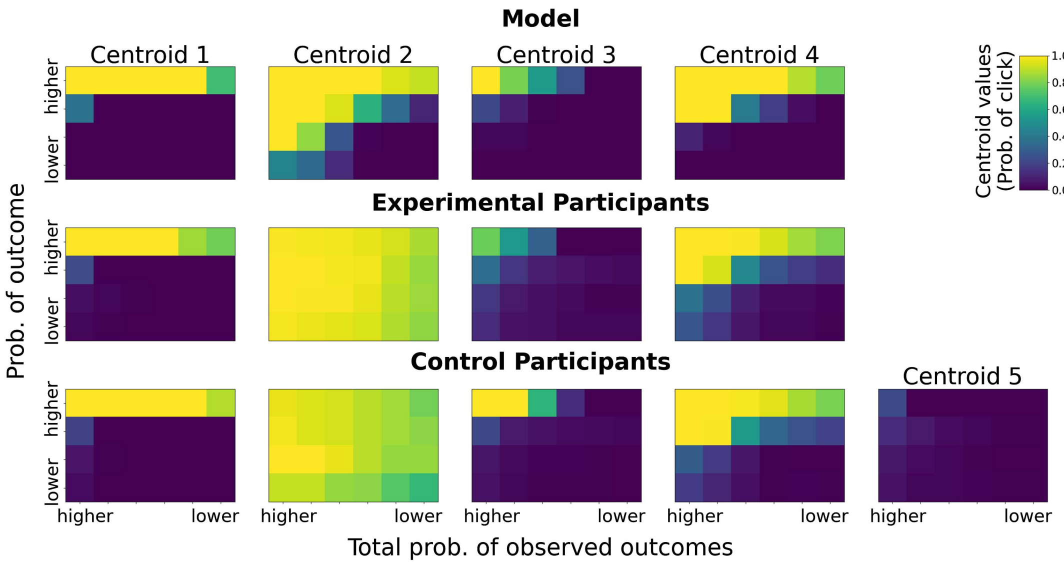
100 Balls	Option 1 value: 39	Option 2 value: -15	Option 3 value: -9	Option 4 value: 43	Option 5 value: -6	Option 6 value: -7
29 Blue	27	-53	-32	-18	75	68
22 Green	29			149		
26 Yellow	9			54	-108	-104
23 Red	98			8		

Total Click Cost: 14 Points

You can bet in 3 seconds

Note. To reduce implicit costs associated with information gathering and information use, participants in the experimental group were given a 20-s time minimum per trial and a display of the subjective expected value of every gamble. See the online article for the color version of this figure.

Figure 10
Experiment 2 k-Means Centroids



Note. The manipulations in the experimental group led to a great reduction in random gambling for participants in this group, which is why a cluster for random gambling was unnecessary (middle panels). The model (top panel) and both groups of participants performed the Mouselab task in low-stakes environments, with a 2 × 2 between-subjects design of outcome dispersion and cost of information gathering. prob. = probability. See the online article for the color version of this figure.

Results

Identification of Strategies

We applied the same *k*-means clustering procedure used in the previous experiment, separately for the model, the experimental group, and the control group. As shown in Figure 10, the clusters in the control group closely matched those found in Experiment 1. However, the experimental group did not contain a distinct cluster for gambling randomly because random gambling was greatly diminished for this group (6.6% vs. 28.6% of all trials, $p < .001$; 4.5% vs. 27.2% of participants gambled randomly on more than half of all trials, $p < .001$). As described in detail below, this brought the strategies of participants in the experimental group into greater alignment with the optimal strategies predicted by our model.[5]

Comparison of Strategies Across Environments

For brevity, we use the following acronyms when referring to the different environments: LD-LC for low dispersion, low cost; LD-HC for low dispersion, high cost; HD-LC for high dispersion, low cost; and HD-HC for high dispersion, high cost.

As illustrated in Figure 11, participants in the experimental group showed an overall shift toward more costly strategies. In all environments, a χ^2 test of independence revealed an increase in the use of targeted search, LD-LC: $\chi^2(1, 3960) = 32.9$, $p < .001$, $d = 0.25$; LD-HC: $\chi^2(1, 3960) = 49.9$, $p < .001$, $d = 0.32$; HD-LC: $\chi^2(1, 3960) = 32.9$, $p < .001$, $d = 0.25$; HD-HC: $\chi^2(1, 3960) = 32.9$, $p < .001$, $d = 0.25$. Conversely, participants used the more frugal SAT-TTB strategy less often in all environments except for the LD-HC environment, LD-LC: $\chi^2(1, 3960) = 12.4$, $p < .001$, $d = -0.16$; LD-HC: $\chi^2(1, 3960) = 0.1$, $p < .8$, $d = 0.01$; HD-LC: $\chi^2(1, 3960) = 12.4$, $p < .001$, $d = -0.16$; HD-HC: $\chi^2(1, 3960) = 12.4$, $p < .001$, $d = -0.16$. While these overall changes away from the frugal SAT-TTB heuristic toward the more costly Target Search strategy brought participants in the experimental condition closer to the predictions of our resource-rational model, they shifted too far toward the most costly strategy, exhaustive search. While the model never uses exhaustive search, participants in the experimental group used it more than those in the control group in all environments except HD-LC, LD-LC: $\chi^2(1, 3960) = 114.9$, $p < .001$, $d = 0.53$; LD-HC: $\chi^2(1, 3960) = 124.6$, $p < .001$, $d = 0.69$; HD-LC: $\chi^2(1, 3960) = 114.9$, $p < .001$, $d = 0.53$; HD-HC: $\chi^2(1, 3960) = 114.9$, $p < .001$, $d = 0.53$. For a detailed comparison of the frequency of each strategy for each group in each environment against our resource-rational model, see Figure 11 (this figure omits random gambling to facilitate comparison with Figure 4; to see the reduction in random gambling in the experimental group, see Figure D5).

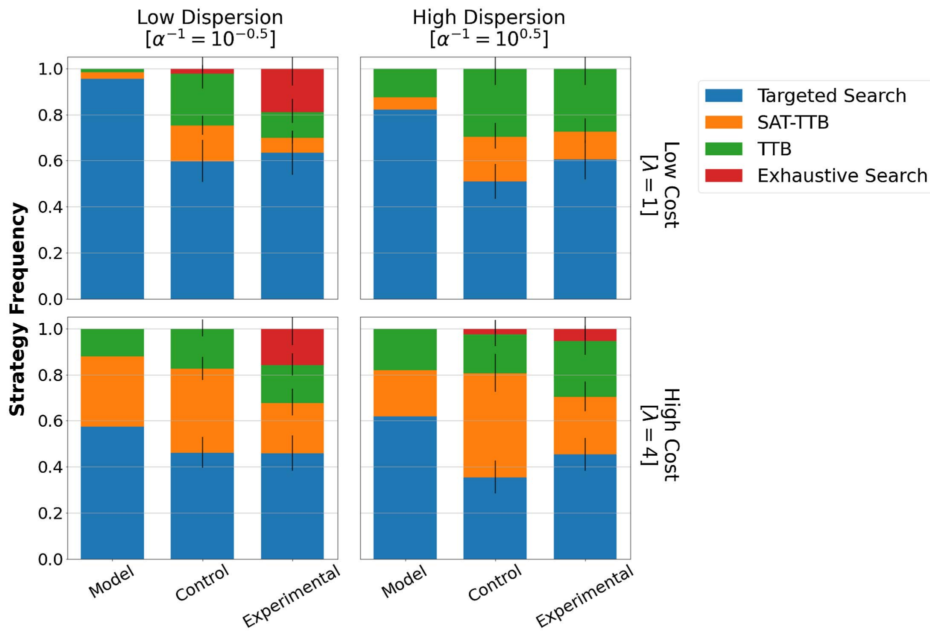
Information Gathering and Choice Behavior

The shift toward more costly strategies in the experimental group is apparent in an overall increase in information gathering compared to the control group, as shown in Figure 12. In each environment, participants in the experimental group gathered more information

[5] The clusters discovered for the model are not identical to those seen in Figure 3, corresponding to Experiment 1, because (a) the environments in Experiment 2 are different; in particular, they are limited to low-stakes environments and do not include any conditions where the cost of gathering information is zero, as in Experiment 1, and (b) the particular trials presented to participants within the low-stakes condition are not identical across experiments (and all model comparisons use the same distribution of trials that are presented to participants).

Figure 11

Reducing Implicit Costs Increases the Use of Costly Strategies



Note. Participants in the experimental group in Experiment 2 show a general increase in the use of targeted search, and even exhaustive search, and a general decrease in the most frugal heuristic, SAT-TTB. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

than those in the control group, two-sample t tests; LD-LC: $t(102) = 4.88$, $p < .001$, $d = 0.96$; LD-HC: $t(100) = 4.71$, $p < .001$, $d = 0.93$; HD-LC: $t(100) = 3.23$, $p < .0017$, $d = 0.64$; HD-HC: $t(94) = 3.31$, $p < .0013$, $d = 0.68$. Participants' levels of information gathering were closer to that of the model than participants in the control group in all environments except LD-HC (Figure 12, top panels). In the LD-HC environment, participants in the experimental group actually gathered *too much* information (Figure 12, bottom panels). These absolute deviations of participant mean information gathering from the model was improved significantly in the experimental group compared to the control group only in the HD-LC condition, LD-LC: $t(102) = -0.38$, $p < .7$, $d = -0.07$; LD-HC: $t(100) = 0.74$, $p < .46$, $d = 0.15$; HD-LC: $t(100) = -2.65$, $p < .0094$, $d = -0.52$; HD-HC: $t(94) = -0.45$, $p < .65$, $d = -0.09$.

We additionally inspected the same three behavioral features of alternative- and attribute-based information processing as in Experiment 1, and these results are presented in Appendix E.

Decision Quality

The model achieved an average decision quality of 0.886; that is, it achieved 88.6% of the unboundedly optimal expected payoff (the difference from Experiment 1 is due to the different environment parameters). Similar to Experiment 1, participants in the control group achieved an average decision quality of 0.479. By contrast, the experimental group achieved an average decision quality of 0.678. Excluding low-effort participants increased these values to .621 for the control group and .788 for the experimental group, with the model's decision quality equals to .879 for the same trials. This suggests that about 44% of the control group's total performance gap relative to unboundedly optimal performance stems from unaccounted cognitive limitations. The resource-rational model explains an additional 32% of this gap. The remaining 24% appear to result from people's deviations from resource-rational decision making.

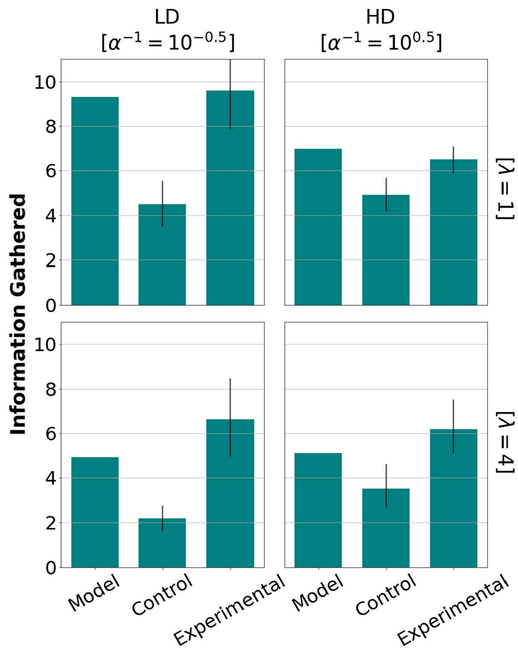
Net Performance

Given that participants in the experimental group behaved more similar to the optimal model in some manners/cases but less similar in others, we next asked how participants' overall (net) performance was affected by the experimental intervention. Recall that this measure includes the cost of clicking as well as the expected payoff of the chosen gamble. Thus, this measure of performance does not give an unfair advantage to participants in the experimental group, who gather more information.

As illustrated in Figure 13, participants in the experimental group achieved numerically higher net performance in three of the four environments, but not in the LD-HC environment. This improvement, however, was only significant in the HD-LC environment, $t(100) = 2.60$, $p < .011$, $d = 0.52$. The difference was not significant in any other environment, LD-LC: $t(102) = 1.51$, $p < .13$, $d = 0.30$; LD-HC: $t(100) = -1.77$, $p < .079$, $d = -0.35$; HD-HC: $t(94) = 0.73$, $p < .47$, $d = 0.15$. Across all conditions, participants in the experimental group were not significantly more resource-rational than participants in the control group, $t(402) = 1.12$, $p < .26$, $d = 0.11$.

Participants in the experimental group should be expected to choose the gamble with the highest subjective expected value on

Figure 12
Information Gathering for Each Group in Experiment 2



Note. Participants in the experimental group successfully increased their information gathering near levels of the model in the low-cost conditions (upper panels) but gathered excessive information in the high-cost conditions (lower panels). Error bars show 95% confidence interval. LD = low dispersion; HD = high dispersion; LC = low cost; HC = high cost. See the online article for the color version of this figure.

100% of trials, since they were given these values (see Figure 9). However, they failed to do so on 17.6% of all trials. As a result, they actually lost *more* points per trial on average than participants in the control group as a result of these errors, 3.3 versus 1.6 points per trial, $t(402) = -2.34$, $p < .02$, $d = -0.23$. This counterintuitive result is manifestly an artifact of participants not performing the task in good faith, since participants in the experimental group were given the best option. Whereas low-effort participants have the option to gamble randomly in the control group or in Experiment 1, in the experimental group, they are forced to wait 20 s. It appears that such low-effort participants gambled randomly after gathering excessive information during the forced wait. To address this, we excluded an equal fraction of participants from both groups based on participant deviation from model performance (see the section on Experiment 2 in Appendix F for details).

When excluding low-effort participants from both groups, participants in the experimental group were significantly more resource-rational than participants in the control group in every condition, LD-LC: $t(55) = 2.36$, $p < .022$, $d = 0.63$; LD-HC: $t(80) = 4.02$, $p < .001$, $d = 0.90$; HD-LC: $t(78) = 2.26$, $p < .027$, $d = 0.51$; HD-HC: $t(73) = 2.30$, $p < .024$, $d = 0.53$.

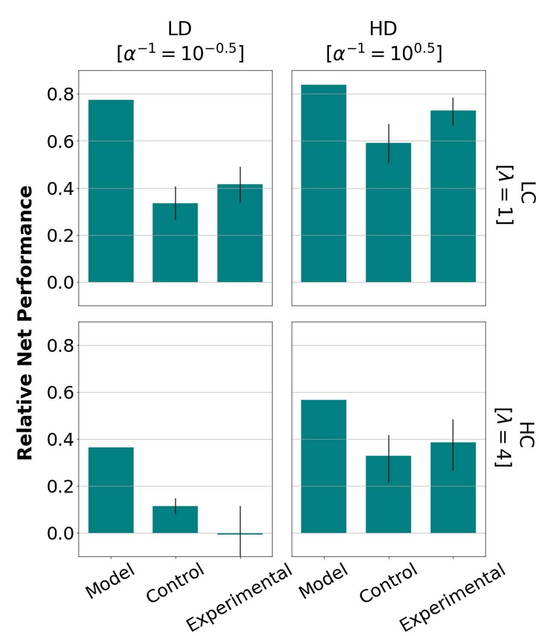
Sources of Underperformance

As in Experiment 1, we measured participants' net performance and four sources of underperformance as a percentage of the model's net performance. Figure 14 compares these results for participants from each condition, showing that participants in the control and experimental groups achieved 55.1% (95% CI [47.8, 59.5]) and 62.1% (95% CI [51.5, 67.6]) of the net performance of the model, respectively. Below we break these gaps down into their four possible sources, as we did in Experiment 1.

Consistent with the goal of our manipulation, participants in the experimental group gathered about the same amount of information as the resource-rational model on average across all environments: The fit implicit cost of clicking was 0.0 points per click for the experimental group and 2.9 for the control group. As a result, performance for participants in the experimental group was not degraded due to implicit costs, while this accounted for a large portion of underperformance for participants in the control group (0.0% vs. 9.6%, respectively, as shown in Figure 14). Surprisingly, participants in the experimental group showed *more* imperfect use of information than participants in the control group (8.9% vs. 4.2%). As described in the previous section, this is due to low-effort participants in the experimental group not performing the task as instructed, since the values were given to make perfect use of information. As shown in Figure F7, when excluding low-performing participants, imperfect information use accounted for 2.1% of underperformance in the control group and 1.9% in the experimental group.

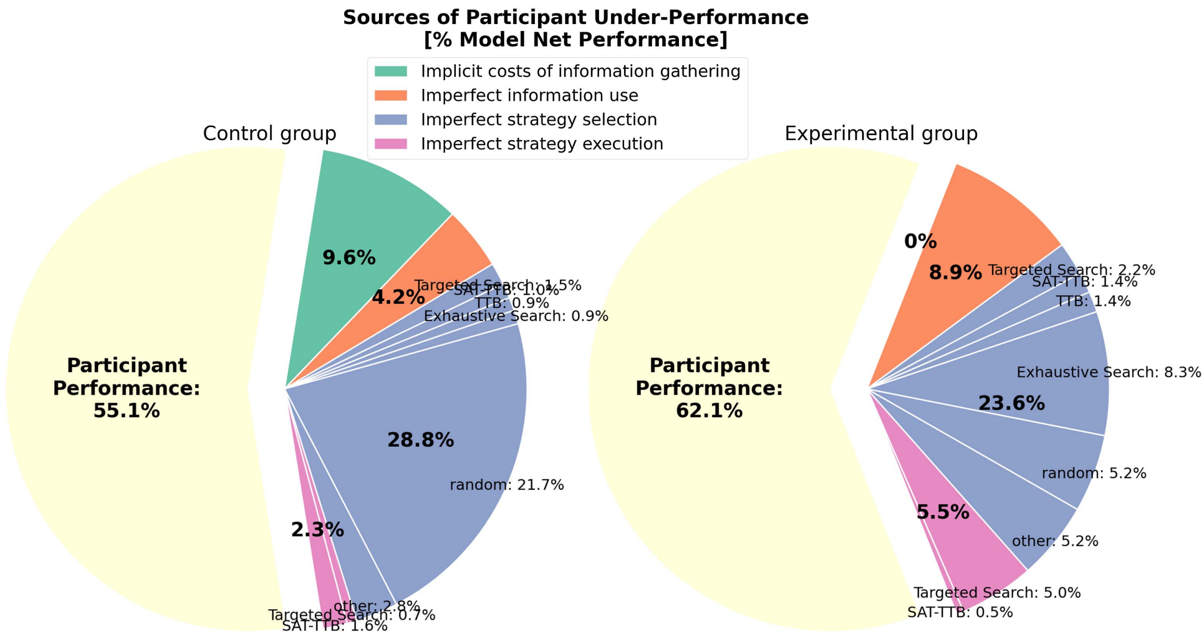
We next considered how imperfect strategy selection and execution differed between the two groups. As shown in Figure 14, imperfect strategy selection accounted for 28.8% of underperformance in the

Figure 13
Net Performance (Including Click Cost) Across Conditions for Each Group in Experiment 2



Note. To facilitate comparison across conditions, we normalize net performance by the expected maximum gamble value. Participants in the experimental condition tended to perform better, but not in all conditions (see Figure F6, for a comparison when excluding low-effort participants). Error bars show 95% confidence interval across participants. LD = low dispersion; HD = high dispersion; LC = low cost; HC = high cost. See the online article for the color version of this figure.

Figure 14
Source of Underperformance for Each Group in Experiment 2



Note. Each pie chart shows the percentage of model net performance achieved by participants in beige, with the remaining percentage (the performance gap) broken up into different sources of underperformance. Compared to participants in the control group (left pie chart), participants in the experimental group (right pie chart) showed slightly better overall performance, with no implicit costs of gathering information, and much less reduction from random gambling. Because some participants in the experimental group do not follow the instructions to make perfect use of information, Figure F7 shows the same results after excluding low-effort participants from both groups. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

control group and 23.6% in the experimental group, while imperfect strategy execution accounted for 2.3% and 5.5%, respectively. Consistent with Experiment 1, engaging in random gambling was the most frequent instance of imperfect strategy selection for participants in the control group, alone accounting for 27.7% of underperformance. In the experimental group, this proportion was reduced to 5.2%. On the other hand, whereas the use of exhaustive search accounted for only 0.9% of underperformance in the control group, it accounted for 8.3% in the experimental group, more than any other strategy. While imperfect execution accounted for a modest proportion of underperformance in both groups, it was slightly more in the experimental group, due to the increased usage of the difficult-to-execute targeted search strategy. Figure 15 shows the sources of imperfect strategy selection (off-diagonal values) and execution (diagonal values) from every strategy. It shows that of the 8.3% reduction in performance from incorrectly selecting exhaustive search in the experimental group, most of it—4.9%—occurred when the best strategy to select was targeted search.

Discussion

The experimental manipulations in Experiment 2 were effective at reducing the implicit cost of information gathering identified in Experiment 2. The most pronounced effect of increased information gathering in the experimental group was a reduction of random gambling and an increase in the use of exhaustive search and targeted search. However, in the high-cost conditions, participants in the experimental group actually gathered too much information. Surprisingly, we did not find that imperfect use of information was reduced in the experimental group (in fact, it increased, although not after excluding low-effort participants). When excluding low-effort participants, we did find that participants in the experimental group were significantly more resource-rational than participants in the control group. However, there was still room for improvement in this group. Overall, these findings suggest that people deviate from resource-rational decision making even in settings where the assumptions of the resource-rational model are met.

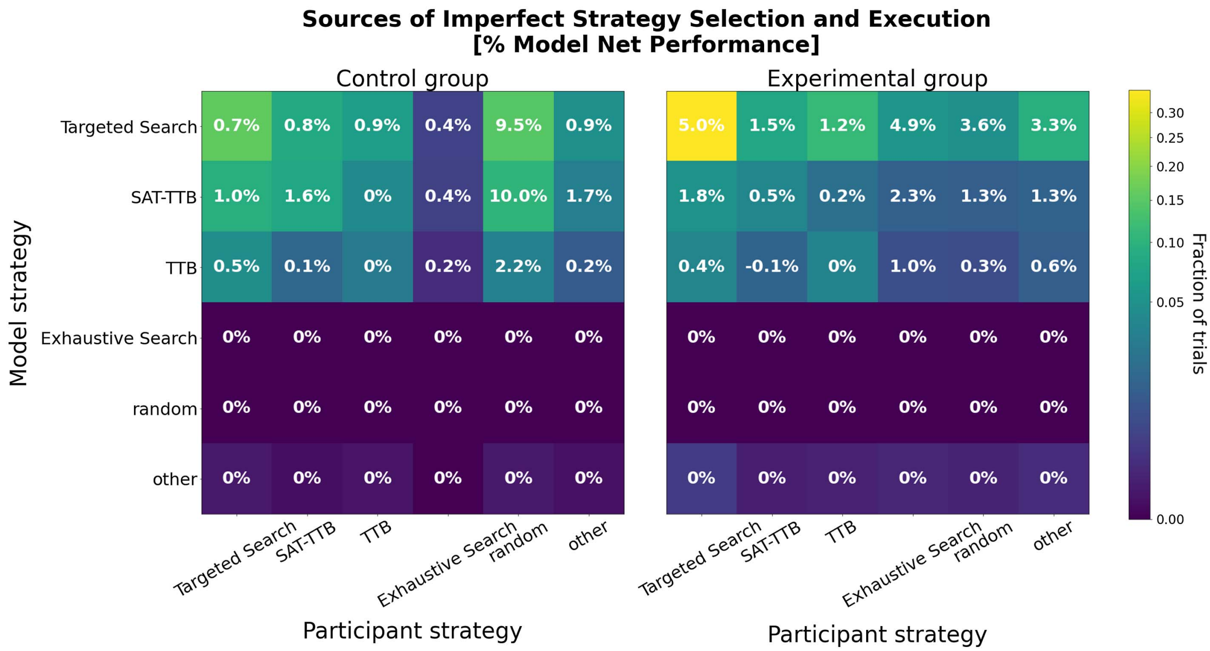
General Discussion

Traditionally, rational models and the heuristics and biases approach have offered very different views of human decision making. As a result, researchers studying human decision making have typically had to choose between assuming people are rational or characterizing their behavior as the result of following heuristics that result in systematic biases. Each approach has advantages and disadvantages. Assuming rationality makes it easy to generate predictions across a wide range of circumstances, but people sometimes systematically deviate from rational principles. Research on heuristics and biases has characterized these deviations, but with many possible heuristics, it can be difficult to predict what people will do in novel situations.

In this work, we have offered a way to reconcile these two perspectives—rationality and heuristics—by deriving optimal

Figure 15

Sources of Imperfect Strategy Selection (Off-Diagonal Values) and Imperfect Strategy Execution (Diagonal Values) for Each Strategy, for the Control Group (Left Plot) and the Experimental Group (Right Plot) in Experiment 2



Note. Experimental participants' excessive use of exhaustive search occurred mostly when they should have used targeted search, according to the model, while control participants' excessive use of random gambling occurred mostly when they should have used SAT-TTB. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

strategies for multi-alternative, multi-attribute decision making from a rational analysis of how people should allocate their limited cognitive resources. This approach of applying rationality to cognitive processes themselves provides a general framework for understanding decision making that can also make task-specific predictions. Drawing on ideas from artificial intelligence and machine learning, we were able to both establish a normative basis for previously identified heuristics and also discover new strategies that had previously been overlooked. Furthermore, we collected a large data set to test our method across a very broad range of decision environments, demonstrating both the generalizability and accuracy of our approach. Our results show that people use all the strategies that our method identified, and they adaptively select which strategy to use in a way that is consistent with our framework. However, the match was by no means perfect; there is still room to improve on human decision making.

One of the key ideas behind our approach is that we can formulate the problem of discovering decision strategies and predicting when they should be used as one of finding the optimal policy of a meta-level MDP (Hay et al., 2012; Russell & Wefald, 1991b). The meta-level MDP framework allows us to identify those decision strategies that optimally trade off the costs associated with acquiring information to update one's beliefs about the world with the benefits of that information. This results in a normative view of heuristics, providing a reconciliation between these historically divergent views of decision making.

Of course, the idea that people trade-off decision accuracy and effort is not new (Beach & Mitchell, 1978; Kool & Botvinick, 2018; Lieder & Griffiths, 2017; Shah & Oppenheimer, 2008; Shenhav et al., 2017). Indeed, intuitions about minimizing effort underlie most work on heuristics, including early work with the Mouselab paradigm (Payne et al., 1988). More recent work has explicitly formalized decision making in the Mouselab paradigm from a resource-rational perspective (Gabaix et al., 2006). The meta-level MDP framework provides a new set of computational tools for understanding heuristics through this lens. In particular, it allows us to identify strategies that strike an optimal trade-off between computational costs and decision quality. Moreover, our approach can be naturally generalized to other tasks.

While the idea that people trade-off decision accuracy and effort is not new (Beach & Mitchell, 1978; Shah & Oppenheimer, 2008), and information gathering in Mouselab has previously been studied from a resource-rational perspective (Gabaix et al., 2006), the meta-level MDP framework provides a new set of computational tools for understanding heuristics through this lens. The result is that we can formally identify information search strategies that achieve an optimal trade-off between computational costs and decision quality. By automatically deriving decision strategies from a normative model, we can avoid the cumbersome and inexact process of searching for heuristics by hand that psychologists have relied on in the past.

In addition to offering a normative standard for evaluating heuristics, the meta-level MDP formalism makes our resource-rational framework generally applicable to any decision-making process. This formalism breaks down decision making into an arbitrary discrete set of cognitive operations and then applies reinforcement learning to this decision-making process itself. This provides a general-purpose approach for deriving optimal heuristics that avoids the need to search an intractable combinatorial space of possible heuristics. It also provides a normative benchmark for evaluating heuristics, that is, by the total meta-level reward they achieve.

We demonstrated the usefulness of this approach using the Mouselab task, which is a classic, well-studied process-tracing paradigm (Payne et al., 1993). While the Mouselab task has been widely used to study decision strategies, these studies are typically limited to around 20–40 participants (e.g., Arieli et al., 2011; Bieleke et al., 2020; Dieckmann & Rieskamp, 2007; Lohse & Johnson, 1996; Payne et al., 1988; Reisen et al., 2008; Rieskamp & Otto, 2006), rarely exceed 100 (Dhar et al., 1999; Mata et al., 2007; Sen, 1999), and the largest study that the authors are aware of collected 255 participants in a 2×2 between-subjects design, which examined the interaction between negative affect and choice difficulty on decision strategies (Stone & Kadous, 1997). In the present study, we searched for heuristics across a broad space of decision environments and tested whether strategies change across the parameters of those environments. This necessitated a large-scale experiment using the Mouselab task. Future work may apply our meta-level MDP framework to potentially any kind of decision-making process, providing a general-purpose, normative approach for understanding how people think and derive strategies for making decisions.

We found that participants used the same four strategies as the resource-rational model; how did they acquire these strategies? It is typically assumed that people have a limited toolkit of general-purpose heuristics that are adapted to real-world environments (e.g., Gigerenzer & Selten, 2002; Hutchinson & Gigerenzer, 2005; Klein, 2008). More specifically, heuristics are thought to develop slowly through evolution and/or learning, rather than being crafted on the fly at decision time. One consequence of this is that, in addition to limitations in cognitive resources and time, humans have a limited toolkit of heuristics to deploy—those which they have previously acquired through evolution and learning (Gigerenzer & Selten, 2002). That these general-purpose heuristics turn out to be resource-rational in our task highlights the effectiveness of these strategies and perhaps the usefulness of the Mouselab task in capturing important characteristics of real-world risky choice.

In addition to offering a method for deriving optimal heuristics, our approach provides a more realistic framework for both evaluating and improving human decision making. To rigorously evaluate and improve decision making, we should understand the agent's computational goal and how it goes about solving it. The resource-rational analysis presented here is an attempt to reverse engineer this decision process by comparing human behavior to the predictions of our resource-rational model. In our experiment, people did indeed use the same strategies as the resource-rational model. Furthermore, the heuristic solutions arising from our framework are inherently sensitive to the statistics of the decision environment—including the stakes of possible reward, the dispersion of possible outcomes, and the cost of acquiring information—and people adapted their strategies to the decision environment in a manner largely consistent with resource-rationality. While participants' performance was consistent with rational use of cognitive resources, they performed below the level of the resource-rational model (Figure 6). Crucially, the underperformance persisted even when we modified the environment in such a way that the assumptions of our resource-rational model were met (Experiment 2). This suggests that human decision making still has room for improvement, even when people's cognitive constraints are taken into account. Our method could be used to provide feedback and teach people which heuristics to use and under what circumstances, in a manner that accounts for their cognitive limitations, providing a computationally informed path to improving human decision making (Becker et al., 2022; Callaway, Jain, et al., 2022; Consul et al., 2022; Mehta et al., 2022; Skirzyński et al., 2021).

Why did people underperform relative to the resource-rational strategies? First, it is important to note that our normative framework should not be mistaken for a descriptive account. Rather, it provides a prescriptive account of how people ought to behave in the Mouselab task. It is therefore not surprising that participants earned less reward than the resource-rational model. Indeed, a key contribution of our approach is that it allowed us to characterize in detail how and (to some extent) why people deviated from the resource-rational benchmark. While these sources of underperformance suggest specific ways that people could improve their decision-making strategies, achieving perfect resource-rationality may still be unattainable. In fact, given that resource-rational decision making is itself an intractable problem (Russell, 2016), this is almost certainly the case. Importantly, however, this does not undermine the value of the approach, for many of the same reasons that traditional rational or "computational level" analyses are useful (Anderson, 2013; Marr, 1982). Providing a rational benchmark for resource-constrained agents reveals both the strengths and weaknesses of human decision making and suggests important directions for future research.

Another possible explanation for the underperformance, one which we did not consider above, is that the computations people use are different from those assumed by our model. Specifically, we assumed an idealized set of cognitive operations based on Bayesian updating, such that each piece of revealed information is perfectly integrated into a posterior belief about the expected payoff of the corresponding gamble. But if that integration process is itself composed of multiple costly operations (e.g., multiplication and addition), then people might not—and indeed, should not—fully integrate all revealed information. This would result in worse performance given the same number of clicks. Applying our method with a fine-grained set of operations (which may include the "EIPs" discussed above) is thus an important direction for future work. This could also be used to account for subjective attitudes toward risk, as discussed above, and other subjective biases (e.g., priors about unrevealed information for unlikely attributes that are biased toward large losses or rewards). These modifications may allow us to better understand the sources of implicit costs of information gathering and the imperfect use of information observed in the current experiments. By expanding the set of computational actions available, we can potentially identify more nuanced strategies and achieve an even closer correspondence to human behavior. Refining the model in this way could also allow it to account for the imperfect strategy selection and imperfect strategy execution we observed relative to the current model.

Limitations

This work is a first proof of concept that meta-level MDPs can be used to understand the decision-making process that gives rise to heuristics. However, as is often the case with computational modeling, the ecological validity of the paradigm and modeling framework comes with a trade-off in computational tractability and precision. Below, we highlight a few such simplifications that we hope will be improved upon in future work.

Unlike usual experiments with the Mouselab paradigm, in this work, probabilities were revealed at no cost. This prevents us from discovering or empirically observing heuristics that do not use probabilities (Glöckner & Pachur, 2012, e.g., minimax, maximax, equal weighting). The main reason for this is that we do not currently have a way to compute the VOI for probabilities. Furthermore, in many situations, revealing probabilities would be more valuable than collecting information, but in the real world, these probabilities do not have to be reasoned out in the same way as the counterfactual benefits of alternative attributes in hypothetical situations (e.g., "How much better would it be to have a comfortable sofa bed when someone visits us?") because they are largely learned from repeated experience (Hertwig et al., 2004, "This happens about thrice a year"). Hence, the real-world applicability of an explicit operation to reveal a probability is arguably limited as well.

Another substantial modification we made to the Mouselab paradigm lies in how information is revealed. In the original version of the Mouselab paradigm, information could be revealed at no cost (besides the time and effort of moving the cursor), and it was only visible while the cursor was hovering over the corresponding cell. This is thought to mimic the real-world process of gathering information through eye movements (Glöckner & Betsch, 2008; Lohse & Johnson, 1996; Reisen et al., 2008, although there is conflicting evidence as to how similar the two processes are). In contrast, in our version of the task, there was an explicit cost for revealing each piece of information, and it remained visible for the remainder of the trial after being revealed. The explicit cost is intended to operationalize the cognitive cost of evaluating an outcome (Bakkour et al., 2019; Biderman et al., 2020, e.g., by memory recall). While these external click costs do not capture all internal cognitive costs (discussed further below), our assumption is that the click costs in our paradigm outweigh purely cognitive costs. As for leaving the information visible, this was purely a pragmatic choice. Because the model does not account for working memory constraints (discussed below), it cannot predict or account for the "reacquisition" of previously revealed information—something that is also true of all the classic heuristic decision strategies we have considered. However, as we discuss below, our resource-rational framework is well-suited to develop such models. Developing such models and evaluating them in the original version of the Mouselab paradigm (where each outcome is only visible while the mouse is hovering over it) is a promising direction for future research.

Perhaps the most substantial limitation of the current work is our simplistic model of cognitive cost. Below we identify three specific ways the cost model could be improved.

First, we assume a uniform and fixed cost of considering any piece of information. In the real world, however, some pieces of information may be more costly to consider than others. For example, some information may be less readily accessible in memory, or even be absent from memory entirely, necessitating external information search. Furthermore, the cost of considering information may vary over time, for example, increasing as the deadline for a decision approaches. Fortunately, it would be relatively easy to account for these types of costs in the model and paradigm, by varying the cost for revealing different cells or increasing the cost over the course of a trial. Accounting for such varying costs in a resource-rational model may uncover new types of rational heuristics that this work could not identify.

Second, the model does not account for the cost of integrating information. The current model treats "consideration" of a feature as an atomic operation, with a fixed cost. However, this operation actually involves several suboperations such as weighing outcomes by probabilities and updating expected values. Critically, some of these operations can be avoided when only one outcome is considered (at a time). Accounting for such costs could provide a rational account of strategies like the lexicographic semiorder heuristic, which do not account for all considered information in its choices. Pursuing such a model will likely require developing more advanced computational methods for approximating the value of computation, as it is not clear how to quantify the VOI produced by these lower level operations.

Third, the model does not account for the cost of maintaining information in working memory. Although such costs are mitigated to some extent by our version of the Mouselab task (with revealed values remaining on screen), expected values that combine information in a column must still be maintained, and this is something our model does not account for. Such costs would be even more important in naturalistic decisions where most or all information must be represented internally. Future work should extend our model to account for these costs, perhaps drawing on previous resource-rational models of working memory allocation and maintenance outside of decision-making contexts (Suchow & Griffiths, 2016; van den Berg & Ma, 2018; Yoo et al., 2018).

Accounting for these additional costs would allow our model to predict not only which pieces of information are acquired but also how they are used in the service of making a choice. This could help the model explain cases where people do not choose the option with the maximum expected value given the revealed information. Furthermore, it would allow our method to discover heuristics like equal weighting and elimination by aspects, which do not use an optimal Bayesian decision rule. Identifying implicit costs that make these kinds of strategies rational is an especially exciting direction for future research.

Measuring the decision rule used subsequent to information search, and how participants use probabilities, would allow for more precise identification of heuristics. In the case of TTB and SAT-TTB, since only a single attribute is considered, there is only one reasonable way to use the revealed information; but in the case of target search and exhaustive search, the final decision rule depends on how probabilities are weighed. Assessing how probabilities are used would allow for the distinction between WADD versus an equally weighted strategy, and whether Target Search corresponds to elimination-by-aspects or another heuristic. We measured strategies primarily through patterns of information search, but some heuristics are defined by additional characteristics. Future work should measure how probabilities are used and the final decision rule used after the information search.

A final limitation of this work is that we have only identified which heuristics are rational to use in different settings (modulo our assumptions about cost), but we have not attempted to explain how people learn those heuristics nor how they select between them. We have merely compared the information search behavior resulting from human learning and strategy selection to the rational use of rational heuristics prescribed by our resource-rational analysis. This comparison suggests that people deploy heuristics adaptively, but not perfectly. That is, just as there are bounds on human rationality in the traditional sense, there appear to also be bounds on human resource-rationality. Future work should explore process-level theories of strategy learning and selection that can explain this pattern of behavior.

Conclusion

Overall, our findings show that participants use resource-rational decision strategies in an adaptive manner, suggesting that people have highly effective mechanisms for discovering and selecting good heuristics. Understanding those mechanisms and how they emerge is an important direction for future research. On the other hand, the deviations from resource-rationality suggest that people might experience additional costs and that their mechanisms for discovering and applying heuristics are imperfect. Future research should attempt to characterize these costs, investigate how people discover heuristics, and develop interventions that improve people's capacity to discover and adaptively choose between heuristics.

References

Analytis, P. P., Kothiyal, A., & Katsikopoulos, K. V. (2014). Multi-attribute utility models as cognitive search engines. *Judgment and Decision Making*, *9*(5), 403–419. https://doi.org/10.1017/S1930297500006781

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429. https://doi.org/10.1037/0033-295X.98.3.409

Anderson, J. R. (2013). *The adaptive character of thought*. Psychology Press.

Arieli, A., Ben-Ami, Y., & Rubinstein, A. (2011). Tracking decision makers under uncertainty. *American Economic Journal: Microeconomics*, *3*(4), 68–76. https://doi.org/10.1257/mic.3.4.68

Bakkour, A., Palombo, D. J., Zylberberg, A., Kang, Y. H. R., Reid, A., Verfaellie, M., Shadlen, M. N., & Shohamy, D. (2019). The hippocampus supports deliberation during value-based decisions. *eLife*, *8*, Article e46080. https://doi.org/10.7554/elife.46080

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multiattribute choice. *Operations Research*, *56*(5), 1289–1304. https://doi.org/10.1287/opre.1070.0485

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, *3*(3), 439–449. https://doi.org/10.5465/amr.1978.4305717

Becker, F., Skirzynski, J., van Opheusden, B., & Lieder, F. (2022). Boosting human decision-making with AI-generated decision aids. *Computational Brain & Behavior*, *5*(3), 467–490. https://doi.org/10.1007/s42113-022-00149-y

Bell, D. E., Raiffa, H., & Tversky, A. (1988). *Decision making: Descriptive, normative, and prescriptive interactions*. Cambridge University Press.

Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., Pinto, H. P. d. O., Raiman, J., Salimans, T., Schlatter, J., … Zhang, S. (2019). *Dota 2 with large scale deep reinforcement learning*. arXiv preprint arXiv:1912.06680. https://doi.org/10.48550/arXiv.1912.06680

Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk. In *Commentaries of the imperial academy of sciences of petropolita* (Vol. *5*, pp. 175–192).

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, *22*(1), 23–36. https://doi.org/10.2307/1909829

Bettman, J. R., Johnson, E. J., & Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes*, *45*(1), 111–139. https://doi.org/10.1016/0749-5978(90)90007-V

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Bhatia, S., & Stewart, N. (2018). Naturalistic multiattribute choice. *Cognition*, *179*, 71–88. https://doi.org/10.1016/j.cognition.2018.05.025

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21. https://doi.org/10.1016/j.cobeha.2021.02.015

Biderman, N., Bakkour, A., & Shohamy, D. (2020). What are memories for? The hippocampus bridges past experience with future decisions. *Trends in Cognitive Sciences*, *24*(7), 542–556. https://doi.org/10.1016/j.tics.2020.04.004

Bieleke, M., Dohmen, D., & Gollwitzer, P. M. (2020). Effects of social value orientation (SVO) and decision mode on controlled information acquisition—A Mouselab perspective. *Journal of Experimental Social Psychology*, *86*, Article 103896. https://doi.org/10.1016/j.jesp.2019.103896

Binz, M., Gershman, S. J., Schulz, E., & Endres, D. (2022). Heuristics from bounded meta-learned inference. *Psychological Review*, *129*(5), 1042–1077. https://doi.org/10.1037/rev0000330

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semi-order. *Organizational Behavior and Human Decision Processes*, *104*(1), 96–112. https://doi.org/10.1016/j.obhdp.2007.02.001

Bossaerts, P., & Murawski, C. (2017). Computational complexity and human decision-making. *Trends in Cognitive Sciences*, *21*(12), 917–929. https://doi.org/10.1016/j.tics.2017.09.005

Bossaerts, P., Yadav, N., & Murawski, C. (2019). Uncertainty and computational complexity. *Philosophical Transactions of the Royal Society B*, *374*(1766), Article 20180138. https://doi.org/10.1098/rstb.2018.0138

Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280. https://doi.org/10.1016/j.cognition.2008.08.011

Brown, S., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology*, *53*(6), 453–462. https://doi.org/10.1016/j.jmp.2009.09.002

Callaway, F., Gul, S., Krueger, P., Griffiths, T. L., & Lieder, F. (2018). Learning to select computations. In A. Globerson & R. Silva (Eds.), *Proceedings of the thirty-fourth conference of uncertainty in artificial intelligence* (pp. 776–785). AUAI Press.

Callaway, F., Jain, Y. R., van Opheusden, B., Das, P., Iwama, G., Gul, S., Krueger, P. M., Becker, F., Griffiths, T. L., & Lieder, F. (2022). Leveraging artificial intelligence to improve people's planning strategies. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(12), Article e2117432119. https://doi.org/10.1073/pnas.2117432119

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLOS Computational Biology*, *17*(3), Article e1008863. https://doi.org/10.1371/journal.pcbi.1008863

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, *6*, 1027. https://doi.org/10.1038/s41562-022-01411-w

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P., Lieder, F., & Griffiths, T. L. (2021). *Human planning as optimal information seeking*. PsyArXiv. https://doi.org/10.31234/osf.io/byaqd

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, *90*(1), 63–86. https://doi.org/10.1016/S0749-5978(02)00508-3

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience*, *27*(2), 371–378. https://doi.org/10.1523/jneurosci.4421-06.2007

Consul, S., Heindrich, L., Stojcheski, J., & Lieder, F. (2022). Improving human decision-making by discovering efficient strategies for hierarchical planning. *Computational Brain & Behavior*, *5*(2), 185–216. https://doi.org/10.1007/s42113-022-00128-3

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). Oxford University Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*(2), 95–106. https://doi.org/10.1037/h0037613

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective and Behavioral Neuroscience*, *8*(4), 429–453. https://doi.org/10.3758/CABN.8.4.429

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/*N* portfolio strategy? *The Review of Financial Studies*, *22*(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

Dhar, R., Nowlis, S. M., & Sherman, S. J. (1999). Comparison effects on preference construction. *Journal of Consumer Research*, *26*(3), 293–306. https://doi.org/10.1086/209564

Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, *35*(7), 1801–1813. https://doi.org/10.3758/bf03193511

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380–473. https://doi.org/10.1037/h0053870

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, *13*(2), 171–192. https://doi.org/10.1016/0030-5073(75)90044-6

Elkan, C. (2003). Using the triangle inequality to accelerate k-means. In T. Fawcett & N. Mishra (Eds.), *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 147–153). Association for the Advancement of Artificial Intelligence Press.

Fishburn, P. C. (1989). Foundations of decision analysis: Along the way. *Management Science*, *35*(4), 387–405. https://doi.org/10.1287/mnsc.35.4.387

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Frank, M. C. (2013). Throwing out the Bayesian baby with the optimal bathwater: Response to Endress (2013). *Cognition*, *128*(3), 417–423. https://doi.org/10.1016/j.cognition.2013.04.010

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, *96*(4), 1043–1068. https://doi.org/10.1257/aer.96.4.1043

Gardner, J. L. (2019). Optimality and heuristics in perceptual neuroscience. *Nature Neuroscience*, *22*(4), 514–523. https://doi.org/10.1038/s41593-019-0340-4

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267–314. https://doi.org/10.1037/0033-295X.96.2.267

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278. https://doi.org/10.1126/science.aac6076

Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. Oxford University Press.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295x.103.4.650

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75–95). Oxford University Press.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT Press.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(Suppl. 3), 15647–15654. https://doi.org/10.1073/pnas.1014269108

Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1055–1075. https://doi.org/10.1037/0278-7393.34.5.1055

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*(1), 21–32. https://doi.org/10.1016/j.cognition.2011.12.002

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. https://doi.org/10.1037/0033-295X.109.1.75

Griffiths, T. L., Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., & Lieder, F. (2019). Doing more with less: Meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences*, *29*, 24–30. https://doi.org/10.1016/j.cobeha.2019.01.005

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. https://doi.org/10.1111/tops.12142

Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science*, *21*(4), 263–268. https://doi.org/10.1177/0963721412447619

Hawkins, G. E., & Heathcote, A. (2021). Racing against the clock: Evidence-based versus time-based decisions. *Psychological Review*, *128*(2), 222–263. https://doi.org/10.1037/rev0000259

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* (pp. 346–355). AUAI Press.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., & Silver, D. (2018). Rainbow: Combining improvements in deep reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1), 3215–3222. https://doi.org/10.1609/aaai.v32i1.11796

Hogarth, R. M., & Karelaia, N. (2005). Ignoring information in binary choice with continuous variables: When is less "more"? *Journal of Mathematical Psychology*, *49*(2), 115–124. https://doi.org/10.1016/j.jmp.2005.01.001

Hogarth, R. M., & Karelaia, N. (2006). "Take the best" and other simple strategies: Why and when they work "well" with binary cues. *Theory and Decision*, *61*(3), 205–249. https://doi.org/10.1007/s11238-006-9000-8

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*(3), 733–758. https://doi.org/10.1037/0033-295x.114.3.733

Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, *11*(1), 63–90. https://doi.org/10.1023/A:1022631118932

Howard, R. A. (1968). The foundations of decision analysis. *IEEE Transactions on Systems Science and Cybernetics*, *4*(3), 211–219. https://doi.org/10.1109/TSSC.1968.300115

Hutchinson, J. M., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, *69*(2), 97–124. https://doi.org/10.1016/j.beproc.2005.02.019

Huygens, C. (1657). *De ratiociniis in ludo aleae* [The Value of all chances in games of fortune]. Ex officinia J. Elsevirii.

Huygens, C. (1714). *Christiani hugenii libellus de ratiociniis in ludo aleae: Or, the value of all chances in games of fortune; cards, dice, wagers, lotteries, &c. mathematically demonstrated*. S. Keimer.

Jarvstad, A., Rushton, S. K., Warren, P. A., & Hahn, U. (2012). Knowing when to move on: Cognitive and perceptual decisions in time. *Psychological Science*, *23*(6), 589–597. https://doi.org/10.1177/0956797611426579

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management Science*, *31*(4), 395–414. https://doi.org/10.1287/mnsc.31.4.395

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice. *Decision Analysis*, *8*(1), 10–29. https://doi.org/10.1287/deca.1100.0191

Katsikopoulos, K. V., & Martignon, L. (2006). Naïve heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, *50*(5), 488–494. https://doi.org/10.1016/j.jmp.2006.06.001

Keeney, R. L., Raiffa, H., & Meyer, R. F. (1993). *Decisions with multiple objectives: Preferences and value trade-offs*. Cambridge University Press.

Kimball, G. E. (1958). A critique of operations research. *Journal of the Washington Academy of Sciences*, *48*(2), 33–37. https://www.jstor.org/stable/24533760

Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*(3), 456–460. https://doi.org/10.1518/001872008X288385

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour*, *2*(12), 899–908. https://doi.org/10.1038/s41562-018-0401-9

Kwisthout, J., Wareham, T., & van Rooij, I. (2011). Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science*, *35*(5), 779–784. https://doi.org/10.1111/j.1551-6709.2011.01182.x

Lee, M. D., & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review*, *11*(2), 343–352. https://doi.org/10.3758/bf03196581

Lee, M. D., Loughlin, N., & Lundberg, I. B. (2002). Applying one reason decision-making: The prioritisation of literature searches. *Australian Journal of Psychology*, *54*(3), 137–143. https://doi.org/10.1080/00049530412331312704

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, *6*(2), 279–311. https://doi.org/10.1111/tops.12086

Lichtenberg, J. M., & Şimşek, Ö. (2017). *Simple regression models* [Conference session]. Proceedings of the NIPS 2016 Workshop on Imperfect Decision Makers: Admitting Real-World Rationality, Barcelona, Spain.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762–794. https://doi.org/10.1037/rev0000075

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, Article e1. https://doi.org/10.1017/s0140525x1900061x

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, *68*(1), 28–43. https://doi.org/10.1006/obhd.1996.0087

Ludvig, E. A., Bellemare, M. G., & Pearson, K. G. (2011). A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. In E. Alonso & E. Mondragón (Eds.), *Computational neuroscience for advancing artificial intelligence: Models, methods and applications* (pp. 111–144). IGI Global.

Manzini, P., & Mariotti, M. (2012). Choice by lexicographic semiorders. *Theoretical Economics*, *7*(1), 1–23. https://doi.org/10.3982/TE679

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman.

Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). Oxford University Press.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, *52*(1), 29–71. https://doi.org/10.1023/A:1015516217425

Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging*, *22*(4), 796–810. https://doi.org/10.1037/0882-7974.22.4.796

Maule, A., & Hodgkinson, G. (2002). Heuristics, biases and strategic decision making. *Psychologist*, *15*(2), 68–71.

Mehta, A., Jain, Y. R., Kemtur, A., Stojcheski, J., Consul, S., Tošic, M., & Lieder, F. (2022). Leveraging machine learning to automatically derive robust decision strategies from imperfect knowledge of the real world. *Computational Brain & Behavior*, *5*, 343–377. https://doi.org/10.1007/s42113-022-00141-6

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Morgenstern, O., & Von Neumann, J. (1953). *Theory of games and economic behavior*. Princeton University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Prentice Hall.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154. https://doi.org/10.1016/j.jmp.2008.12.005

Nowozin, S. (2014). *Optimal decisions from probabilistic models: The intersection-over-union case* [Conference session]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, United States.

Papadimitriou, C. H., & Tsitsiklis, J. (1986). Intractable problems in control theory. *SIAM Journal on Control and Optimization*, *24*(4), 639–654. https://doi.org/10.1137/0324038

Parpart, P., Jones, M., & Love, B. C. (2018). Heuristics as bayesian inference under extreme priors. *Cognitive Psychology*, *102*, 127–144. https://doi.org/10.1016/j.cogpsych.2017.11.006

Payne, J. W. (1976a). *Heuristic search processes in decision making*. ACR North American Advances.

Payne, J. W. (1976b). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational*

Behavior and Human Performance, *16*(2), 366–387. https://doi.org/10.1016/0030-5073(76)90022-2

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. https://doi.org/10.1037/0278-7393.14.3.534

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. https://doi.org/10.48550/arXiv.1201.0490

Puterman, M. L. (2014). *Markov decision processes: Discrete stochastic dynamic programming*. Wiley.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1226–1243. https://doi.org/10.1037/a0036801

Reisen, N., Hoffrage, U., & Mast, F. W. (2008). Identifying decision strategies in a consumer choice situation. *Judgment and Decision Making*, *3*(8), 641–658. https://doi.org/10.1017/S1930297500001595

Rescorla, R. A., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*(2), 207–236. https://doi.org/10.1037/0096-3445.135.2.207

RStudio Team. (2019). *RStudio: Integrated development environment for R*. http://www.rstudio.com/

Russell, S. (2016). Rationality and intelligence: A brief update. In V. C. Müller (Ed.), *Fundamental issues of artificial intelligence* (pp. 1–21). Springer.

Russell, S., & Wefald, E. (1991a). *Do the right thing: Studies in limited rationality*. MIT Press.

Russell, S., & Wefald, E. (1991b). Principles of metareasoning. *Artificial Intelligence*, *49*(1–3), 361–395. https://doi.org/10.1016/0004-3702(91)90015-C

Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 676–696. https://doi.org/10.1037//0278-7393.9.4.676

Safarzadeh, S., & Rasti-Barzoki, M. (2018). A modified lexicographic semi-order model using the best–worst method. *Journal of Decision Systems*, *27*(2), 78–91. https://doi.org/10.1080/12460125.2018.1498046

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical Association*, *46*(253), 55–67. https://doi.org/10.2307/2280094

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, *31*(3), 699–714. https://doi.org/10.1177/001316447103100310

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and statistical modeling with Python* [Conference session]. 9th Python in Science Conference, Austin, TX, United States. https://doi.org/10.25080/Majora-92bf1922-011

Sen, S. (1999). The effects of brand name suggestiveness and decision goal on the development of brand knowledge. *Journal of Consumer Psychology*, *8*(4), 431–455. https://doi.org/10.1207/s15327663jcp0804_04

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shteingart, H., & Loewenstein, Y. (2014). Reinforcement learning and human behavior. *Current Opinion in Neurobiology*, *25*, 93–98. https://doi.org/10.1016/j.conb.2013.12.004

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354–359. https://doi.org/10.1038/nature24270

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. https://doi.org/10.1037/h0042769

Simon, H. A. (1972). Theories of bounded rationality. In C. B. McGuire & R. Radner (Eds.), *Decision and organization* (pp. 161–176). Elsevier.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*(1), 1–20. https://doi.org/10.1146/annurev.ps.41.020190.000245

Şimşek, Ö. (2013). Linear decision rule as aspiration for simple decision heuristics. In C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 26* (pp. 2904–2912). Curran Associates, Inc.

Şimşek, Ö., & Buckmann, M. (2015). Learning from small samples: An analysis of simple decision heuristics. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems, 28* (pp. 3159–3167). Curran Associates, Inc.

Skirzyński, J., Becker, F., & Lieder, F. (2021). Automatic discovery of interpretable planning strategies. *Machine Learning*, *110*, 2641–2683. https://doi.org/10.1007/s10994-021-05963-2

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, *69*(3), 213–225. https://doi.org/10.1086/258464

Stone, D. N., & Kadous, K. (1997). The joint effects of task-related negative affect and task difficulty in multiattribute choice. *Organizational Behavior and Human Decision Processes*, *70*(2), 159–174. https://doi.org/10.1006/obhd.1997.2703

Suchow, J. W., & Griffiths, T. L. (2016). Deciding to remember: Memory maintenance as a Markov decision process. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2063–2068). The Cognitive Science Society.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497–537). MIT Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, *23*(1), 86–112. https://doi.org/10.1016/0030-5073(79)90048-5

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*(3), 219–225. https://doi.org/10.1002/bs.3830250306

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*(1), 31–48. https://doi.org/10.1037/h0026750

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299. https://doi.org/10.1037/h0032955

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *eLife*, *7*, Article e34963. https://doi.org/10.7554/elife.34963

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … the SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Wübben, M., & Wangenheim, F. v. (2008). Instant customer base analysis: Managerial heuristics often "get it right". *Journal of Marketing*, *72*(3), 82–93. https://doi.org/10.1509/jmkg.72.3.082

Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, *8*, Article 16162. https://doi.org/10.1038/s41598-018-34282-1

Zanakis, S. H., Solomon, A., Wishart, N., & Dublish, S. (1998). Multi-attribute decision making: A simulation comparison of select methods. *European Journal of Operational Research*, *107*(3), 507–529. https://doi.org/10.1016/S0377-2217(97)00147-1

Zednik, C., & Jäkel, F. (2016). Bayesian reverse-engineering considered as a research strategy for cognitive science. *Synthese*, *193*(12), 3951–3985. https://doi.org/10.1007/s11229-016-1180-3

Appendix A

Bayesian Meta-Level Policy Search

Bayesian meta-level policy search (BMPS) is a reinforcement learning algorithm for solving meta-level MDPs that we recently developed to address the computational challenges of strategy discovery (Callaway et al., 2018). BMPS rests on the idea that the value of computation can be approximated by interpolating between the myopic value of computation, the value of perfect information about the gamble that the computation is reasoning about, and the value of perfect information. Concretely, the BMPS policy is defined as

$$\pi_{\text{meta}}(b) = \arg\max_c w_1 \cdot \text{VOI}_1(b, c) + w_2 \cdot \text{VPI}_{\text{sub}}(b, c) + w_3 \cdot \text{VPI}(b) - w_4 \cdot \text{cost}(c), \tag{A1}$$

subject to the constraints that $w_1, \ldots, w_3 \in [0, 1]$, $w_1 + w_2 + w_3 = 1$, and $w_4 > 0$. BMPS identifies a set of weights that maximize the expected return (total meta-level reward) of this policy.

To compute optimal risky choice strategies, we applied BMPS to the meta-level MDP model of decision making in the Mouselab paradigm described in the main text. To achieve this, we instantiated the four features that BMPS uses to approximate the value of computation as follows: First, the value of perfect information is the expected improvement in decision quality if one knew the exact values of every gamble, rather than deciding based on the current belief state. Formally, it is

$$\text{VPI}(b_t) = \mathop{\mathbb{E}}_{v_g^* \sim b_t}\left[\max_g v_g^*\right] - \max_g b_{t,g}^{(\mu)}, \tag{A2}$$

where the expectation over the true gamble values, v_g^*, is taken with respect to the current belief state, capturing the fact that previous computation informs how valuable future computation will be (e.g., if one gamble is already almost certainly better than the others, there is little value to computing more).

Second, the myopic value of information is the expected improvement in decision quality if one executes one more computation before making a decision. Formally, it is

$$\text{VOI}_1(b_t, c) = \mathop{\mathbb{E}}_{b_{t+1} \sim T_{\text{mean}}(b_t, c, \cdot)}\left[\max_g b_{t+1,g}^{(\mu)}\right] - \max_g b_{t,g}^{(\mu)}. \tag{A3}$$

The previous two features provide upper and lower bounds on the true value of executing a computation, based on upper and lower bounds on the amount of future computation that could be executed. We can also consider the value of intermediate amounts of computation; in particular, we use the value of learning the exact value of just one gamble, the one that the considered computation is reasoning about. This is defined as the expected maximum of the true value of that gamble and the current expected value of the best alternative gamble. Formally,

$$\text{VPI}_{\text{sub}}(b_t, c) = \mathop{\mathbb{E}}_{v_{gc}^* | b_{t,gc}}\left[\max\left\{v_{gc}^*, \max_{g \neq gc} b_{t,g}^{(\mu)}\right\}\right] - \max_g b_{t,g}^{(\mu)}, \tag{A4}$$

where g_c is the gamble that computation c is reasoning about and $v_{g_c}^*$ is the (hypothetical) true value of that gamble. As before the expectation is taken with respect to the current belief about the value of the gamble, and we subtract the value of deciding immediately.

Finally, the cost of computation feature was simply

$$\text{cost}(c) = -r_{\text{meta}}(\cdot, c) = \lambda. \tag{A5}$$

We applied BMPS separately to each of the 50 meta-level MDPs modeling the 50 types of decision environments used in the experiment. For each environment, we ran 500 iterations of Bayesian optimization. In each iteration, the algorithm chooses a candidate weight vector and estimates the performance of the corresponding policy averaged across 10,000 simulated decisions. Each of the 10,000 decisions is made in an environment with independent payoff values and outcome probabilities (sampled according to the environment's α and σ parameters). The algorithm then returns the weight vector with the highest expected performance. See Callaway et al. (2018) for details of the BMPS optimization procedure.

(*Appendices continue*)

Appendix B

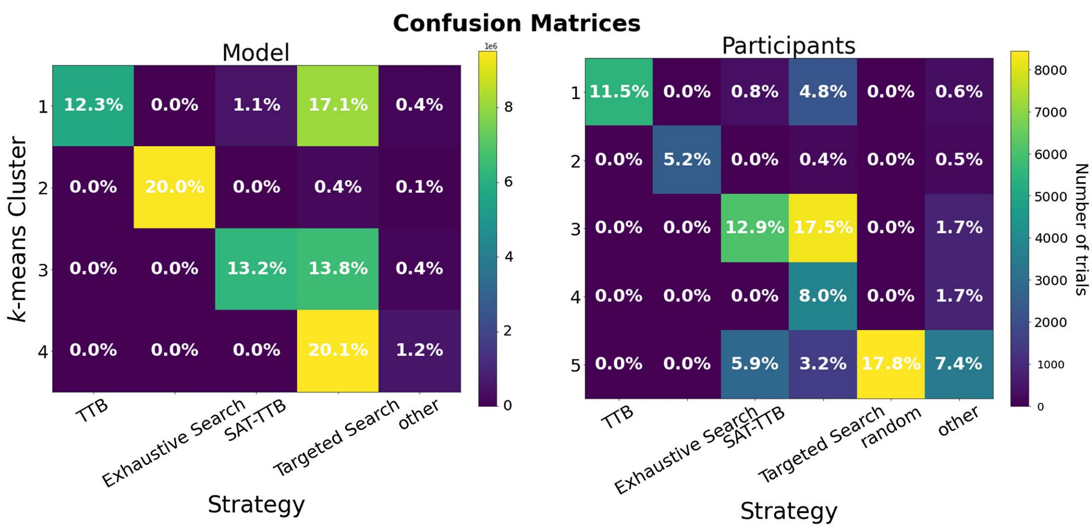
Identification of Resource-Rational Decision Strategies

We took a data-driven approach to discovering heuristic click sequences by applying the *k*-means clustering algorithm to vectors of click sequences. Here, we show the correspondence between cluster labels and heuristic strategies, which are independently defined.

We used $k = 4$ clusters for the model and $k = 5$ for participants, to account for the large portion of random gambling in participants, which does not occur in the model. Here, we show centroids from running *k*-means clustering with values of *k* ranging from 1 to 12.

Figure B1

Confusion Matrices Showing Agreement Between k-Means Cluster Labels and Strategy Definitions for the Resource-Rational Model (Left) and Participant Trials (Right) in Experiment 1

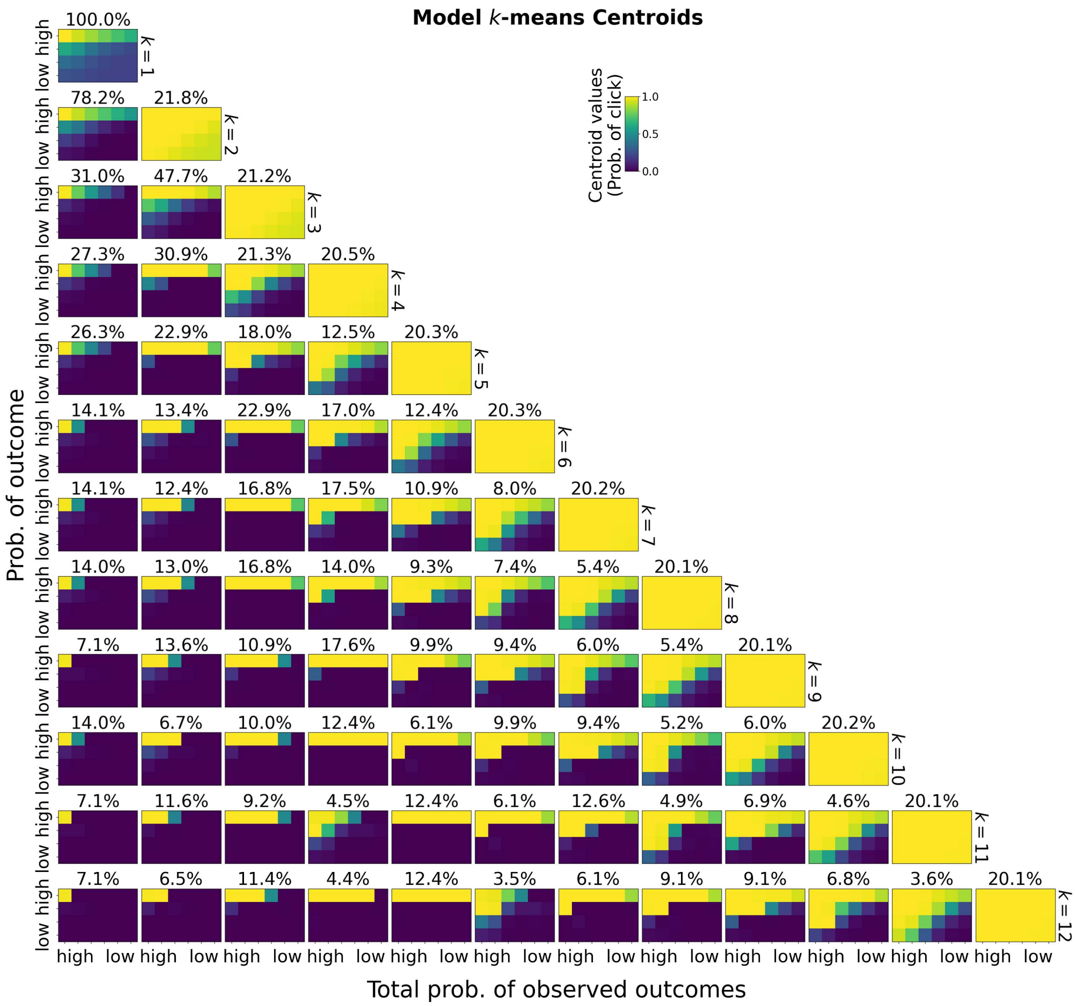


Note. Annotations show the percentage of total trials accounted for by each strategy pair, with colors indicating the trial count. Cohen's $\kappa = 0.572$, 95% CI [0.571, 0.572] for the model, and $\kappa = 0.572$, 95% CI [0.571, 0.572] for participants. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

(*Appendices continue*)

Figure B2

k-Means Clustering Results for Model Data in Experiment 1

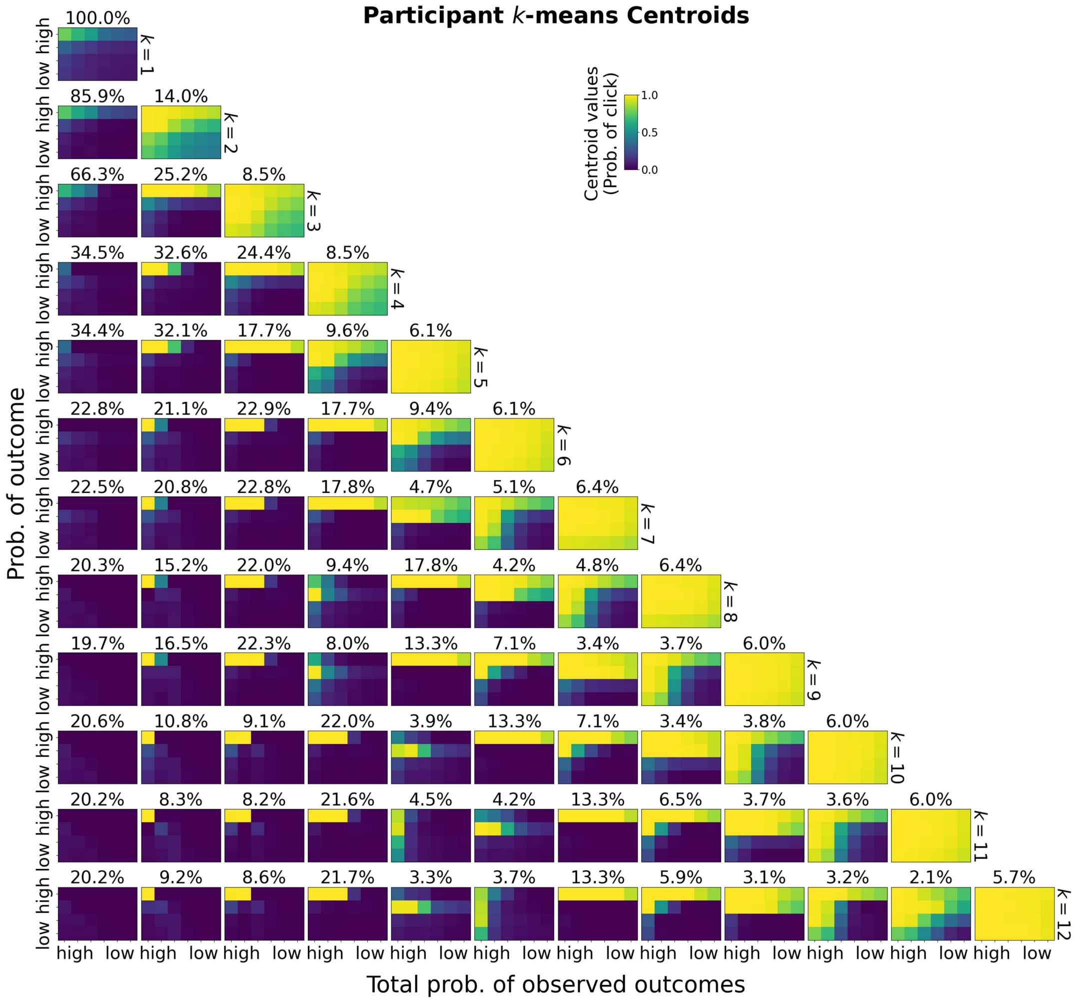


Note. Each row shows the cluster centroid(s) with a number of clusters, k, ranging from 1 to 12. Columns are organized by least to most average information gathering (clicks) per cluster, with subplot titles indicating the percentage of all trial vectors belonging to that cluster. After $k = 4$ clusters, the centroid patterns become largely redundant. prob. = probability. See the online article for the color version of this figure.

(*Appendices continue*)

Figure B3

Same as the Previous Figure but With Participant Data From Experiment 1



Note. prob. = probability. See the online article for the color version of this figure.

Appendix C

Trial-Level Analysis

To assess the extent to which individual participants' information-gathering behavior corresponds to the model in a more fine-grained way, we measured the similarity between participant and model click patterns on individual trials. For each participant trial, we computed the number of clicks that differed between the participant's click vector (identical to that used for *k*-means clustering, as described previously) and the model's click vectors for the same trial. As described previously, we ran 1,000 model simulations for each participant trial since the model may display somewhat different behavior for the same trial. Therefore, for each participant trial, we took the average difference between the participant click vector and all 1,000 model click vectors for that trial. We then averaged this trial-specific difference across all 20 experiment trials for each participant. We then conducted a permutation test to see how many participants showed trial-level behavior that was significantly sensitive to the environment/problem structure in the manner predicted by the model. For each participant, we constructed a null-hypothesis distribution by sampling 1,000 "permuted" differences, computing the average difference in the same way but permuting the model's click vectors such that each participant vector was matched to a model vector from some other trial (which could have been seen by a different participant). We found that 66.8% of participants were significantly ($p < .05$) more consistent with the trial-matched model than with the permuted (trial-mismatched) model.

(*Appendices continue*)

Appendix D

Comparison of Strategies Across Environments

This appendix provides additional details to accompany the sections titled Comparison of Strategies Across Environments for each experiment.

Experiment 1

We inspected how participants adapted their strategy use frequency to the structure of the environment. Figure 4 shows the main effect of each of the three parameters of the environment (stakes, dispersion, and cost) on strategy use frequency for the model and participants; the figures in this section show strategy use frequencies in all 50 environments (with Two Levels of Stakes × Five Levels of Dispersion × Five Levels of Cost). They illustrate the overall qualitative correspondence between the model and participants in the adaptive application of strategies according to the statistics of the environment.

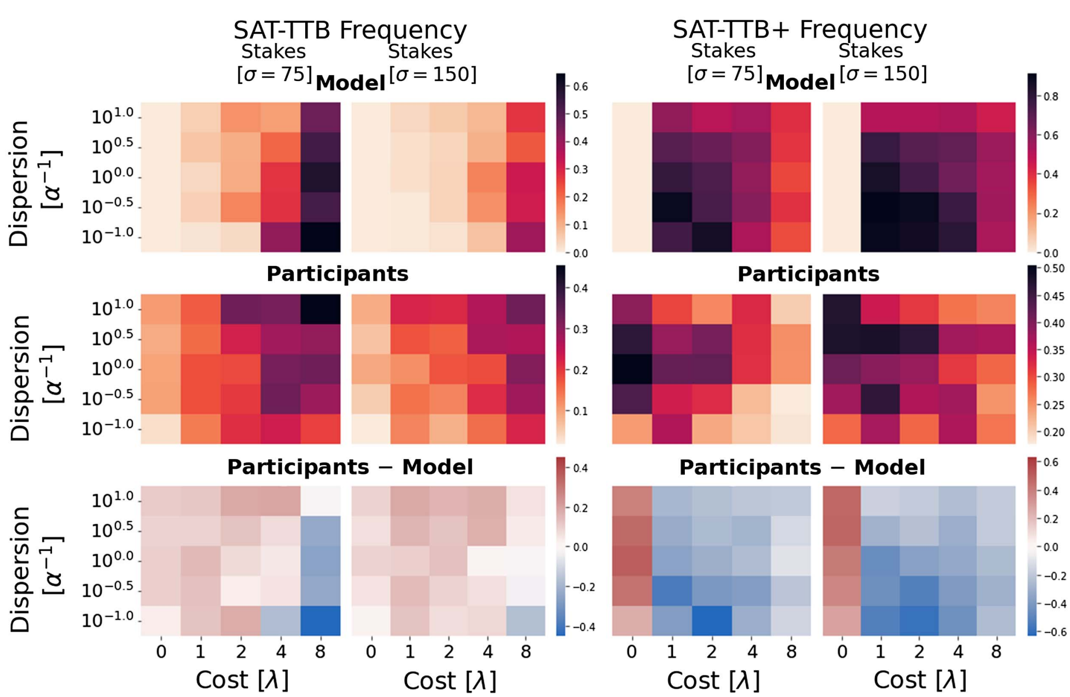
Figure 4 in the main text shows the four main strategies identified, omitting random gambling and other strategies and other, unidentified patterns of clicking. Figures D3 and D4 below show strategy frequencies that include random gambling, and both random gambling and unidentified patterns of clicking, respectively.

Experiment 2

To facilitate comparison with Experiment 1 (Figure 4) in the main text, Figure 11 is conditioned on the same four strategies (i.e., omitting random gambling and unidentified patterns of clicking). Figure D5 includes random gambling to illustrate how much this decreased in the experimental group compared to the control group. Figure D6 includes all trials.

Figure D1

Frequency of SAT-TTB (Left Panels) and Targeted Search (Right Panels) Across All 50 Experimental Conditions, for the Model (Top Panels), Participants (Middle Panels), and a Comparison Between the Model and Participants (Bottom Panels) From Experiment 1

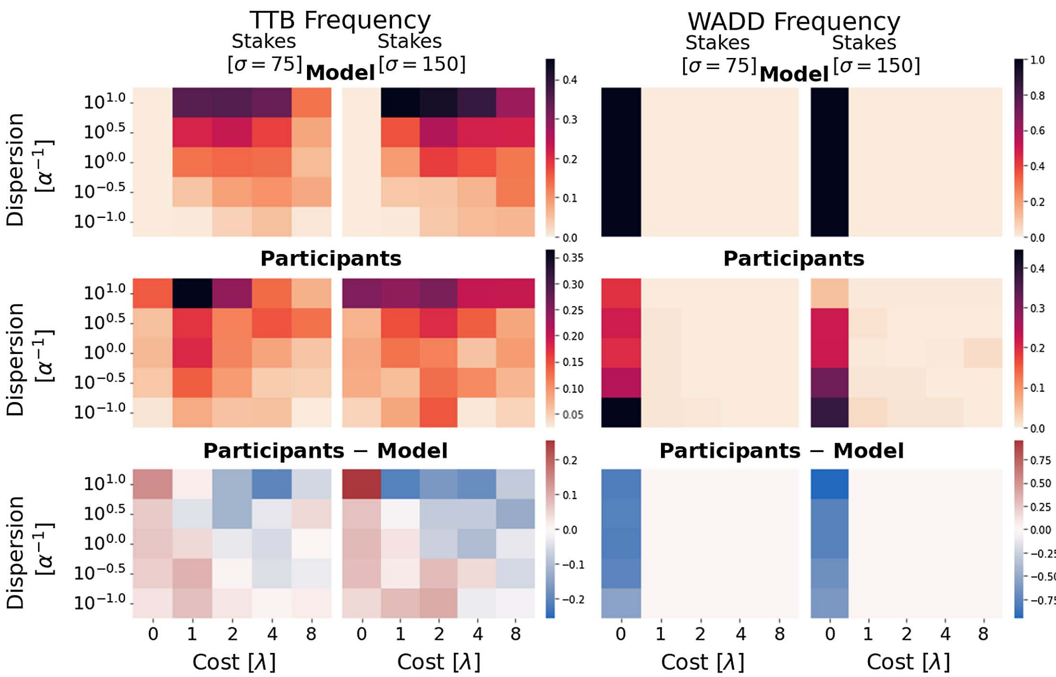


Note. The decision environment in each condition is defined by three parameters: σ (variance in potential reward received), α^{-1} (homogeneity of the outcome distribution), and λ (number of points deducted for each piece of information gathered). The results here accompany the results shown in Figure 4. Targeted search and SAT-TTB are two strategies discovered using our resource-rational method. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

(*Appendices continue*)

Figure D2

TTB (Left Panels) and Exhaustive Search (Right Panels) Strategy Use Frequencies Across All 50 Conditions in the Experiment, for the Model (Top Panels), Participants (Middle Panels), and a Comparison Between the Model and Participants (Bottom Panels) From Experiment 1



Note. TTB and WADD (similar to exhaustive search) are two known heuristics that our resource-rational model rediscovered. The decision environment in each condition is defined by three parameters: σ (variance in potential reward received), α^{-1} (homogeneity of the outcome distribution), and λ (number of points deducted for each piece of information gathered). This figure corresponds to Figure 4, which shows frequencies for each parameter, collapsed across all others. TTB = take-the-best; WADD = weighted additive. See the online article for the color version of this figure.

Table D1

Statistical Results Accompanying Figure 4 From Experiment 1

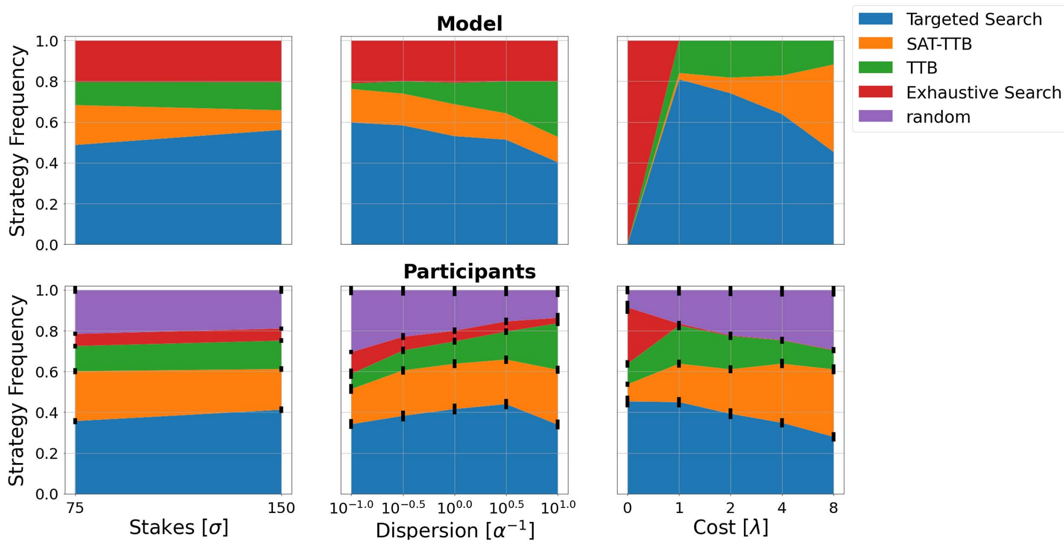
Strategy	Independent variable	Significant post hoc comparison	Effect size (Cohen's *d*)
SAT-TTB	Stakes	n/a	0.11
SAT-TTB+	Stakes	n/a	−0.09
TTB	Dispersion	All pairs	−0.089, −0.048, −0.083, −0.23
Random	Dispersion	All pairs	0.12, 0.051, 0.11, 0.037
SAT-TTB+	Cost	All pairs	0.045, 0.084, 0.078, 0.13
TTB	Cost	All pairs except 0 and 8	−0.21, 0.047, 0.13, 0.063
SAT-TTB	Cost	All pairs	−0.27, −0.078, −0.16, −0.089

Note. Summary of statistical results accompanying the analyses reported in the section Comparison of Strategies Across Environments and shown in Figure 4 from Experiment 1. When applicable, post hoc pairwise comparisons were conducted between all 10 pairs of levels of each independent variable using the Benjamini–Hochberg false discovery rate procedure. This test was not applicable (n/a) when the independent variable had only two levels. The effect sizes for these comparisons were calculated using Cohen's *d* and are presented in ascending order of the corresponding levels of the independent variable (reporting adjacent pairs only). TTB = take-the-best; SAT-TTB = satisficing-TTB.

(*Appendices continue*)

Figure D3

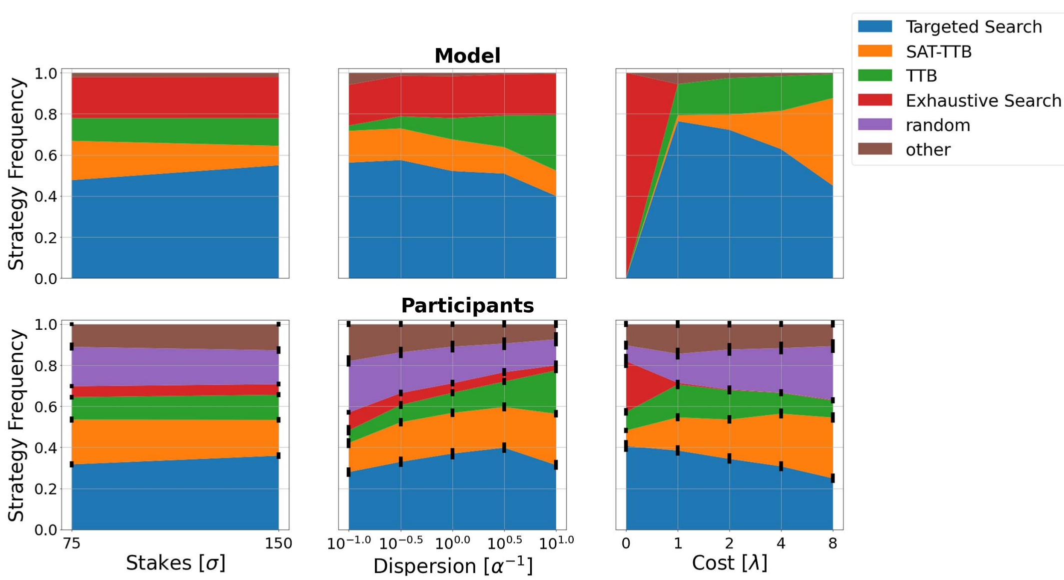
Same as Figure 4 in the Main Text, but Including Trials in Which No Information Was Gathered



Note. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

Figure D4

Same as the Previous Figure but Also Including Unidentified Patterns of Clicking

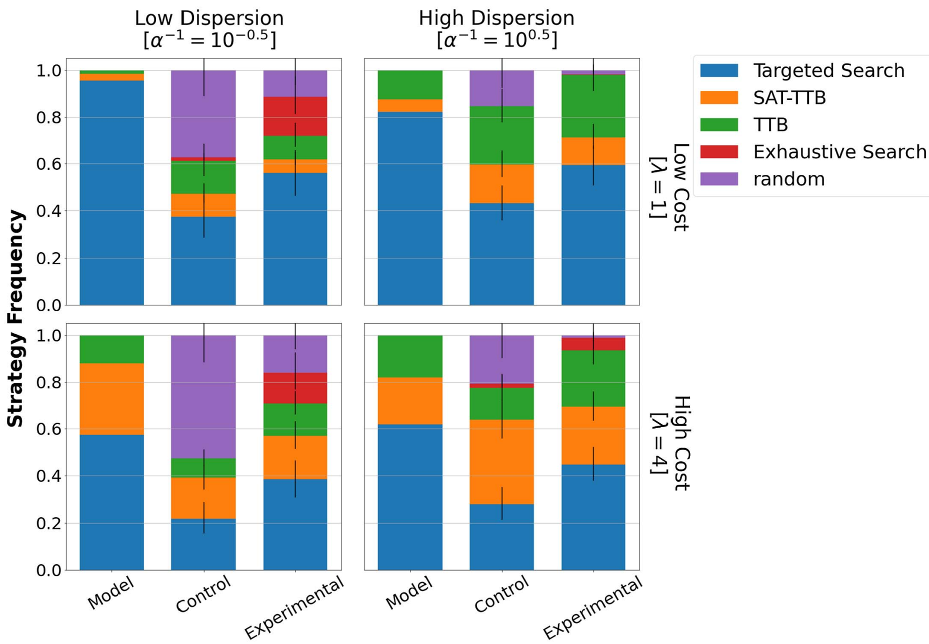


Note. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

(*Appendices continue*)

Figure D5

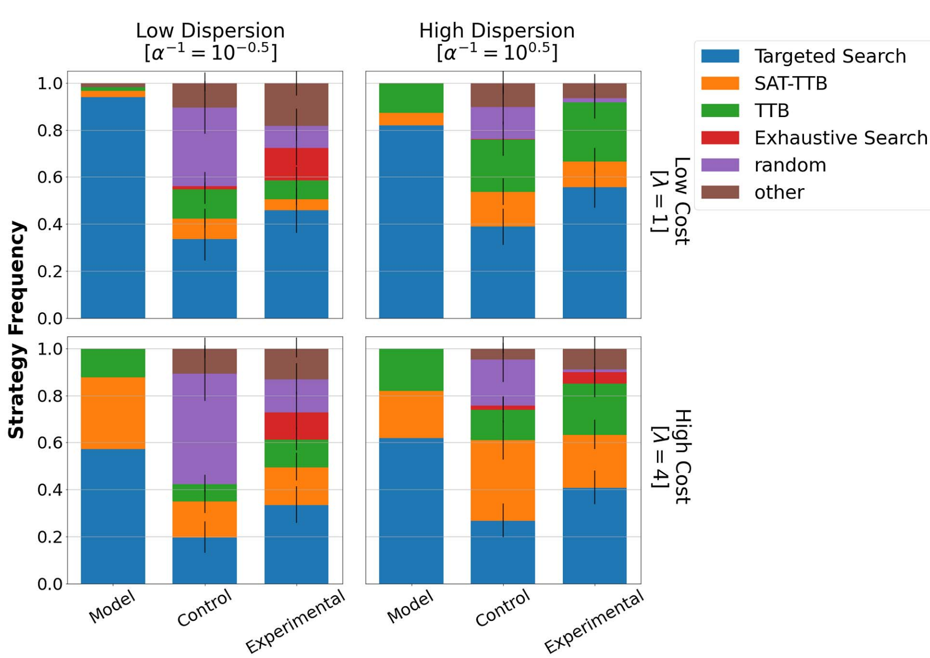
Reducing Implicit Costs Increases the Use of Costly Strategies and Reduces Random Gambling for Participants in the Experimental Group in Experiment 2



Note. Compare with Figure 11, which omits random gambling. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

Figure D6

Same as the Previous Figure but Also Including Unidentified Patterns of Clicking From Experiment 2



Note. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

(*Appendices continue*)

Appendix E

Rational Strategy Selection Explains Variability in Choice Behavior

This appendix provides additional figures and statistical results to accompany the sections Understanding Variability in Choice Behavior for Experiment 1 and Information Gathering and Choice Behavior for Experiment 2.

Experiment 1

Having shown that human participants use the same strategies as the resource-rational model, and adapt them to the environment in much the same way as the model, we next tested theoretical predictions about how four different behavioral characteristics ought to vary with the structure of the environment. The first two are the amount of information gathered and the relative frequency of alternative- versus attribute-based processing. Figure 5 displays the main effect of each of the three parameters of the decision environment on each of these variables. Figure E1 displays these two variables in all 50 environmental conditions. Figures E2 and E3 show the alternative variance and attribute variance. In all cases, participants show a correspondence to the theoretical predictions of the model as to how these behavioral markers should adapt to the environment. See the Rational Strategy Selection Explains Variability in Choice Behavior subsection in the Results section of the main text for details on how these measurements were defined.

Table E1 summarizes statistical analyses accompanying those presented in the main text, corresponding to Figures 5, 6, and E2. A two-sample *t* test was used to calculate the effect of stakes on the dependent variables. One-way analyses of variance were run to assess the effects of dispersion and cost. Post hoc pairwise comparisons were conducted between all 10 pairs of levels of each independent variable using two-sample *t* tests with the Tukey's honestly significant difference correction for multiple comparisons. The effect sizes for these comparisons were calculated using Cohen's *d*.

Experiment 2

In Experiment 2, we inspected the same three information processing features as in Experiment 1: relative alternative- versus attribute-based processing, attribute variance in information gathering, and alternative variance in information processing. As shown in Figure E4, the pattern of results reflects the overall increase in information gathering in the experimental group: decreased attribute variance and alternative variance (Figure E4B and E4C) and less relative emphasis on attribute processing over alternative processing (which is a result of collecting more information since there are more alternatives than attributes; Figure E4A). The statistical results of comparing these measurements across the experimental group and the control group are summarized in Table E2, and similar results comparing these measurements between the model and each group are presented in Table E3, showing that for some measures, the behavior of participants in the experimental group became more similar to the model than that of the control group.

High Dispersion Leads to Attribute-Based Processing

Outcome dispersion is an important determinant of information gathering and strategy selection, with high dispersion favoring attribute-based processing since one attribute is much more likely than others. Figure 12 shows that information gathering decreases with dispersion for the experimental group but increases with dispersion for the control group, and these contrasting patterns can be seen clearly in Figure E6A. We performed a follow-up exploratory analysis to see if this pattern is consistent with the model. The model does indeed predict a two-way interaction between dispersion and cost on information gathering, whereby information gathering decreases with dispersion at low cost but increases with dispersion at high costs. This makes sense intuitively: When the cost of clicking is low, then lower dispersion merits more clicking since the most likely attribute is less informative on average, but when the cost of clicking is high, then higher dispersion allows more frugal clicking that focuses on the most likely attribute. As predicted by the model, when the cost of clicking is low, participants in the experimental group click more with low dispersion, $t(99) = 3.19$, $p = .0019$, $d = .63$, but unlike the model, for high cost, participants in this group click slightly *less* with high dispersion, $t(95) = 0.40$, $p = .69$, $d = 0.08$. The opposite pattern holds for the control group: clicking increases with dispersion for both high cost, as predicted by the model, $t(99) = 2.32$, $p = .022$, $d =$ 46, and low cost, unlike the model, $t(103) = 0.63$, $p = .53$, $d = 0.12$.

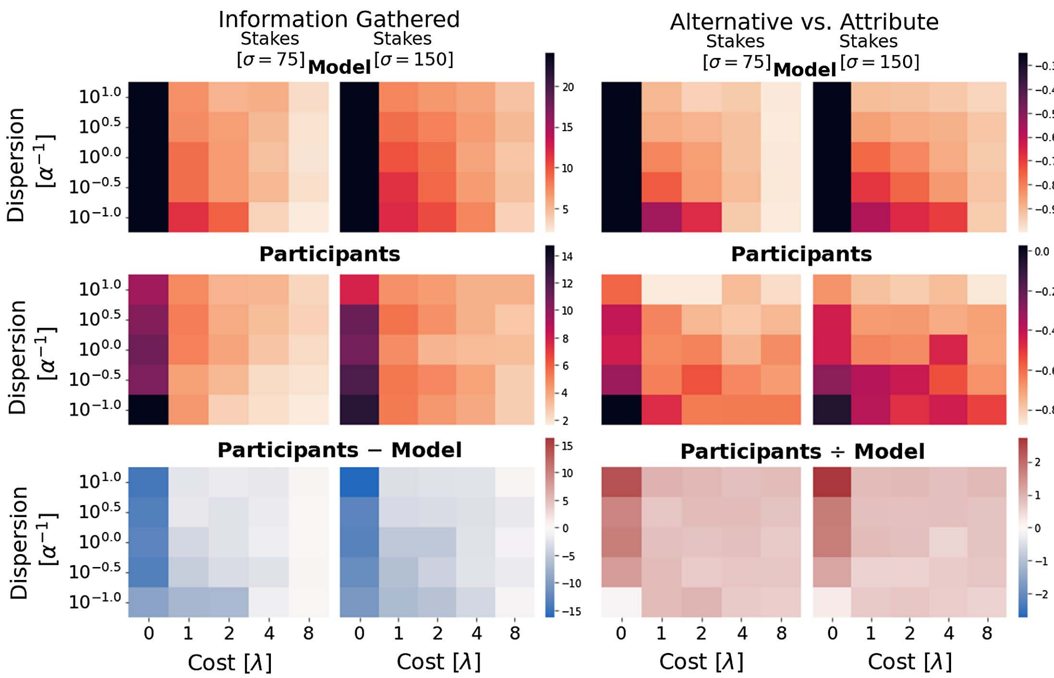
In both groups, these seemingly contradictory results are, in fact, consistent with participants moving toward single attribute-based processing as dispersion increases (as in TTB, which gathers exactly six samples of information, corresponding to the dashed line in Figure E6A). For participants in the experimental group who gather too much information at high cost, information gathering ought to decrease with dispersion, whereas for participants in the control group who gather too little information at low cost, information gathering ought to increase with dispersion. These same predictions can be tested using data from Experiment 1, with five levels of dispersion and cost. As shown in Figure E6B, both the model and participants do indeed display the predicted pattern of results: information gathering shifts toward single attribute processing as dispersion increases, regardless of cost. Rather, the absolute level of information gathering (around six clicks, dashed line) determines the point of reversal in the two-way interaction between dispersion and cost on information gathering. Figure E5 illustrates the same results on a 3D surface.

The same interaction between dispersion and cost for each group in Experiment 2 can be observed for strategy frequencies (Figure 11) and information processing patterns (Figure E4). While participants tend to underperform due to, in part, too little information gathering (in the control group) or too much information gathering (in the experimental group), the overall pattern of how they adapt their information processing to dispersion and cost is broadly consistent with the model's predictions.

(*Appendices continue*)

Figure E1

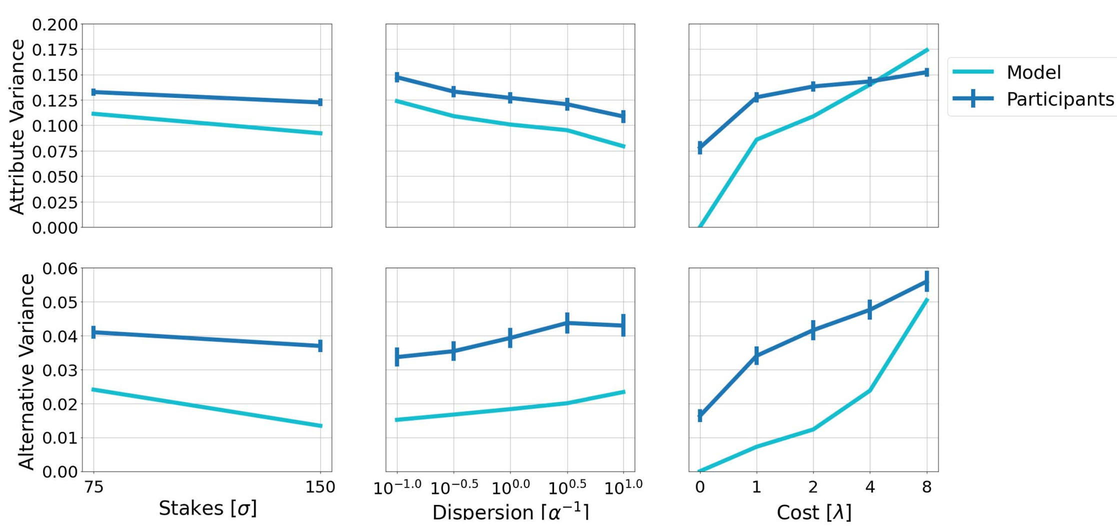
Information-Gathering (Measured With Clicks; Left Panels) and Attribute- Versus Alternative-Based Processing (Right Panels) Shown Across All 50 Conditions of Experiment 1, for the Model (Top Row), Human Participants (Middle Row), and a Comparison Between the Model and Participants (Bottom Row)



Note. The 50 conditions vary in three parameters for a 2 × 5 × 5 across-participant design: reward stakes (σ), uniformity of outcome probabilities (α^{-1}), and the cost per click (λ). The results here accompany the behavioral results shown in Figure 5. Within each parameter value in Figure 5, results are averaged across all values of other parameters, whereas in this figure, the full results for each of the 50 conditions are shown. See the Rational Strategy Selection Explains Variability in Choice Behavior subsection in the Results section of the main text for details on how alternative- versus attribute-based processing was measured. See the online article for the color version of this figure.

Figure E2

Behavioral Correspondence Between Participants and the Resource-Rational Model From Experiment 1

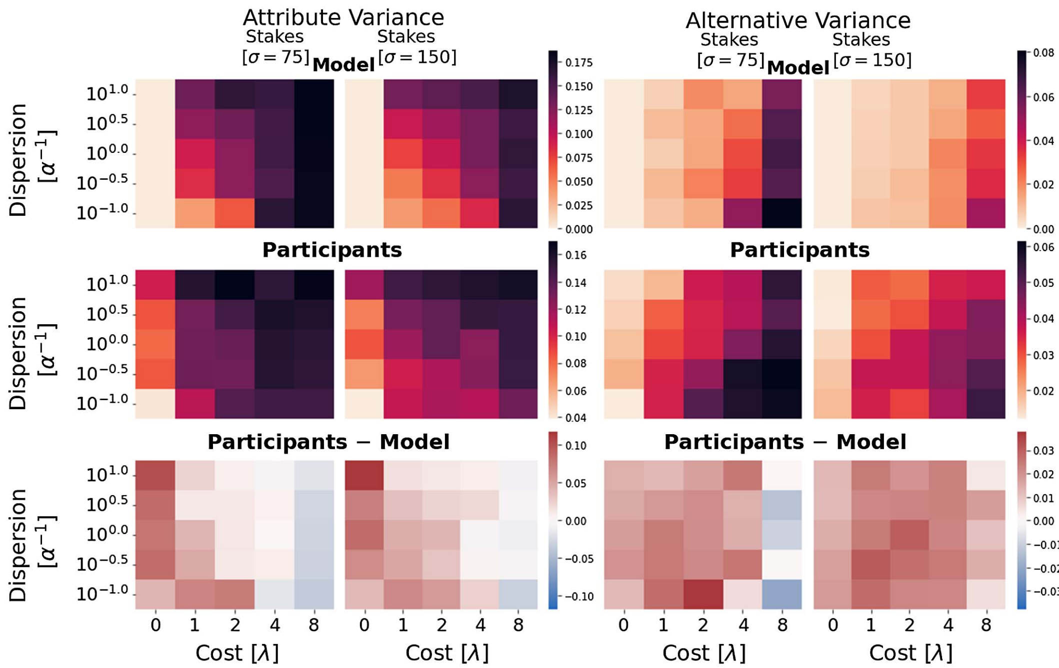


Note. Attribute variance (top panels) and alternative variance (bottom panels) for the resource-rational model and human participants vary across the three parameters of the experiment: σ (reward stakes), α^{-1} (dispersion of outcome probabilities), and λ (cost per click). Error bars show the 95% confidence interval across participants. See the online article for the color version of this figure.

(*Appendices continue*)

Figure E3

Alternative and Attribute Variances for All 50 Conditions in From Experiment 1 (All Combinations of σ, α^{-1}, and λ), for the Model (Top Panels), Participants (Middle Panels), and Difference Between the Two (Bottom Panels)



Note. The results here accompany the behavioral results shown in Figure 5. Within each parameter value in Figure 5, results are averaged across all values of other parameters, whereas in this figure, the full results for each of the 50 conditions are shown. See the online article for the color version of this figure.

Table E1

Statistical Results Accompanying Figures 5 and E2 From Experiment 1

Behavioral feature	Independent variable	Main effect	Significant post hoc comparison	Effect size (Cohen's *d*)
Information gathering	Stakes	$t(2366) = -2.61$, $p = .009$	n/a	−0.11
Information gathering	Dispersion	$F(4, 2363) = 1.22$, $p = .3$	n/a	0.064, 0.0012, −0.036, 0.11
Information gathering	Cost	$F(4, 2363) = 293.83$, $p < .001$	All pairs except 2 and 4, 4 and 8	1.0, 0.32, 0.25, 0.29
Alternative versus attribute	Stakes	$t(2131) = -2.28$, $p = .022$	n/a	−0.099
Alternative versus attribute	Dispersion	$F(4, 2128) = 27.97$, $p < .001$	All pairs except $10^{-1.0}$ and $10^{-0.5}$, $10^{0.0}$ and $10^{0.5}$	0.16, 0.2, 0.092, 0.23
Alternative versus attribute	Cost	$F(4, 2128) = 31.44$, $p < .001$	0 and 1, 0 and 2, 0 and 4, 0 and 8	0.52, 0.048, −0.012, 0.12
Attribute variance	Stakes	$t(2195) = 3.89$, $p < .001$	n/a	0.17
Attribute variance	Dispersion	$F(4, 2192) = 24.74$, $p < .001$	All pairs except $10^{-0.5}$ and $10^{0.0}$, $10^{0.0}$ and $10^{0.5}$	−0.18, −0.1, −0.11, −0.26
Attribute variance	Cost	$F(4, 2192) = 121.75$, $p < .001$	All pairs except 2 and 4, 4 and 8	−0.78, −0.2, −0.095, −0.19
Alternative variance	Stakes	$t(2195) = 2.93$, $p = .0035$	n/a	0.12
Alternative variance	Dispersion	$F(4, 2192) = 8.43$, $p < .001$	$10^{-1.0}$ and $10^{0.5}$, $10^{-1.0}$ and $10^{1.0}$, $10^{-0.5}$ and $10^{0.5}$, $10^{-0.5}$ and $10^{1.0}$	−0.023, 0.14, 0.13, 0.057
Alternative variance	Cost	$F(4, 2192) = 115.01$, $p < .001$	All pairs	−0.7, −0.24, −0.19, −0.26

Note. Summary of statistical results corresponding to the analyses shown in Figures 5 and E2 from Experiment 1. A two-sample *t* test was used to test the main effect of stakes on the dependent variables. Analyses of variance were used to assess the main effects of dispersion and cost. When applicable, post hoc pairwise comparisons were conducted between all 10 pairs of levels of each independent variable using two-sample *t* tests with the Tukey's honestly significant difference correction for multiple comparisons. These tests were not applicable (n/a) when the independent variable had only two levels or its main effect was not significant. The effect sizes for these comparisons were calculated using Cohen's *d* and are presented in ascending order of the corresponding levels of the independent variable (reporting adjacent pairs only).

(*Appendices continue*)

Table E2
Statistical Results Accompanying Figure E4 From Experiment 2

Behavioral feature	Condition (dispersion, cost)	t statistic	p value	Effect size (Cohen's d)
Processing pattern	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 1$	$t(90) = 1.39$	$p = .17$	$d = 0.29$
Processing pattern	$\alpha^{-1} = 10^{0.5}$, $\lambda = 1$	$t(95) = -1.37$	$p = .17$	$d = -0.28$
Processing pattern	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 4$	$t(85) = 2.94$	$p = .0042$	$d = 0.64$
Processing pattern	$\alpha^{-1} = 10^{0.5}$, $\lambda = 4$	$t(86) = 3.43$	$p = .001$	$d = 0.73$
Attribute variance	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 1$	$t(92) = -3.23$	$p = .0017$	$d = -0.67$
Attribute variance	$\alpha^{-1} = 10^{0.5}$, $\lambda = 1$	$t(96) = -0.94$	$p = .35$	$d = -0.19$
Attribute variance	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 4$	$t(88) = -3.16$	$p = .0021$	$d = -0.67$
Attribute variance	$\alpha^{-1} = 10^{0.5}$, $\lambda = 4$	$t(90) = -3.73$	$p = .001$	$d = -0.78$
Alternative variance	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 1$	$t(92) = -2.24$	$p = .027$	$d = -0.46$
Alternative variance	$\alpha^{-1} = 10^{0.5}$, $\lambda = 1$	$t(96) = -2.02$	$p = .046$	$d = -0.41$
Alternative variance	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 4$	$t(88) = -2.21$	$p = .03$	$d = -0.47$
Alternative variance	$\alpha^{-1} = 10^{0.5}$, $\lambda = 4$	$t(90) = -3.54$	$p = .001$	$d = -0.74$

Note. Summary of comparisons between the experimental group and the control group for the behavioral measures shown in Figure E4 from Experiment 2.

Table E3
Statistical Results Accompanying Figure E4 From Experiment 2

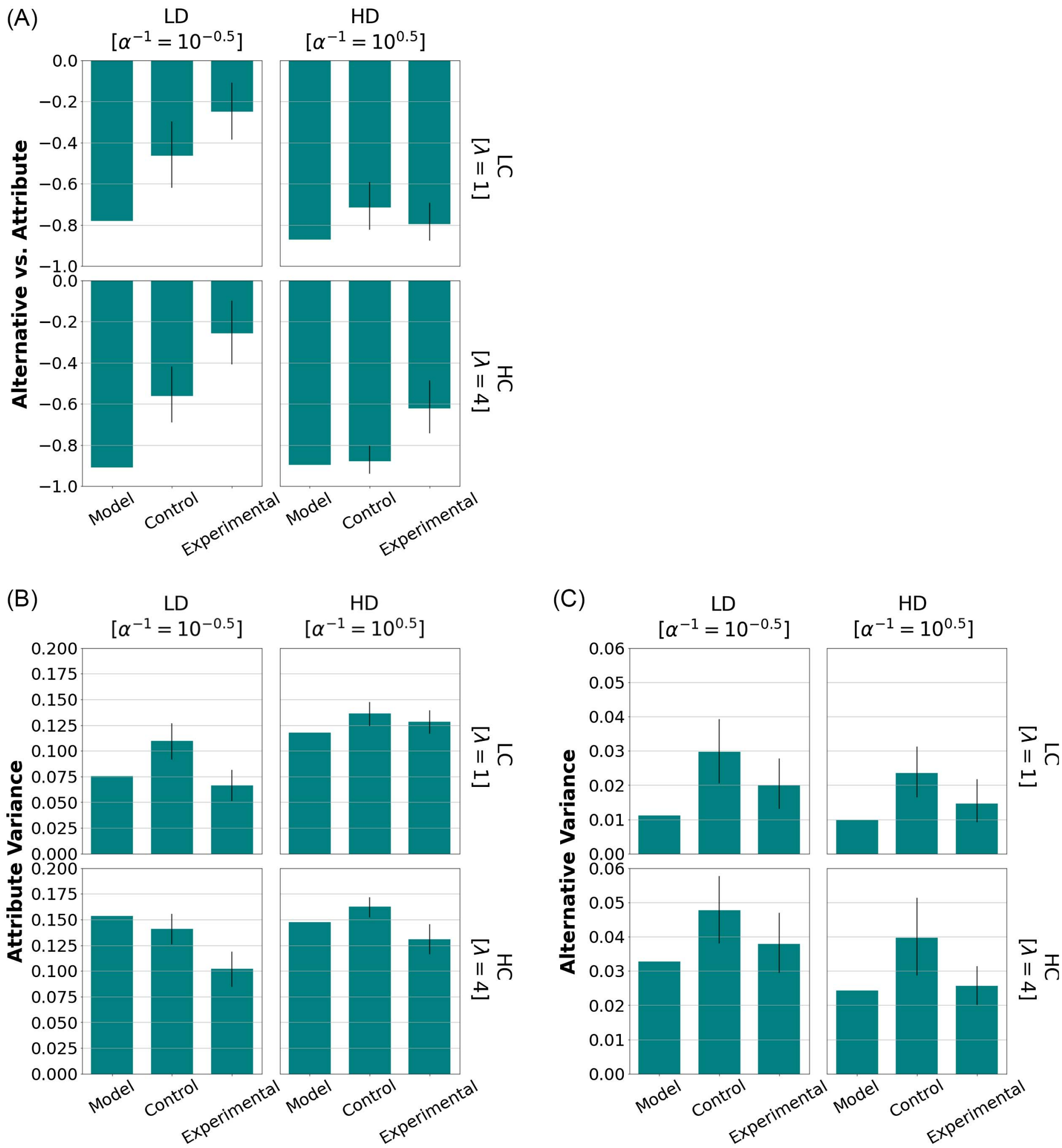
Behavioral feature	Condition (dispersion, cost)	t statistic	p value	Effect size (Cohen's d)
Processing pattern	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 1$	$t(90) = 0.94$	$p = .35$	$d = 0.20$
Processing pattern	$\alpha^{-1} = 10^{0.5}$, $\lambda = 1$	$t(95) = -1.50$	$p = .14$	$d = -0.31$
Processing pattern	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 4$	$t(85) = 2.74$	$p = .0076$	$d = 0.59$
Processing pattern	$\alpha^{-1} = 10^{0.5}$, $\lambda = 4$	$t(86) = 2.91$	$p = .0046$	$d = 0.62$
Attribute variance	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 1$	$t(92) = -1.25$	$p = .21$	$d = -0.26$
Attribute variance	$\alpha^{-1} = 10^{0.5}$, $\lambda = 1$	$t(96) = -0.56$	$p = .58$	$d = -0.11$
Attribute variance	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 4$	$t(88) = 2.83$	$p = .0058$	$d = 0.60$
Attribute variance	$\alpha^{-1} = 10^{0.5}$, $\lambda = 4$	$t(90) = 1.30$	$p = .2$	$d = 0.27$
Alternative variance	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 1$	$t(92) = -1.81$	$p = .073$	$d = -0.37$
Alternative variance	$\alpha^{-1} = 10^{0.5}$, $\lambda = 1$	$t(96) = -2.08$	$p = .04$	$d = -0.42$
Alternative variance	$\alpha^{-1} = 10^{-0.5}$, $\lambda = 4$	$t(88) = 0.03$	$p = .97$	$d = 0.01$
Alternative variance	$\alpha^{-1} = 10^{0.5}$, $\lambda = 4$	$t(90) = -3.11$	$p = .0025$	$d = -0.65$

Note. Summary of comparisons between each group and the model for the behavioral measures shown in Figure E4 from Experiment 2. That is, these statistics report the comparison between groups of each group's absolute deviation from the model, for each dependent variable.

(*Appendices continue*)

Figure E4

Behavioral Features of Information Processing From Experiment 2

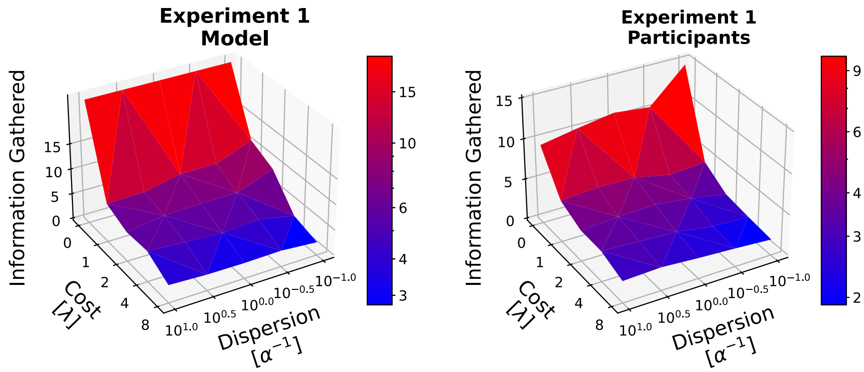


Note. (A) Consistent with their overuse of exhaustive search, participants in the experimental condition showed an increase in alternative versus attribute processing (with negative values indicating relatively more attribute-based processing). (B) Participants in the experimental group showed less overall variance in attribute processing, indicating more use of compensatory strategies that focus on multiple attributes. (C) The same participants showed decreased alternative variance, consistent with increased information gathering evenly across alternatives. LD = low dispersion; HD = high dispersion; LC = low cost; HC = high cost. See the online article for the color version of this figure.

(*Appendices continue*)

Figure E5

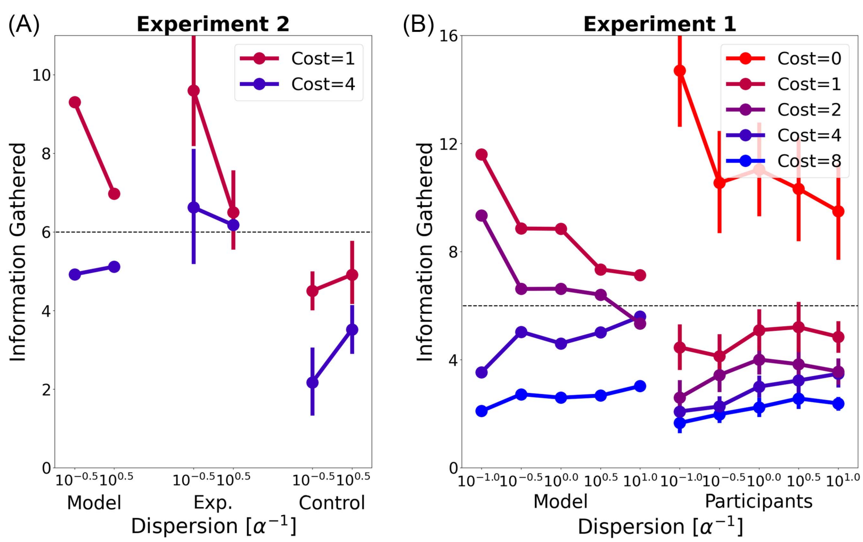
Interaction Between Cost and Dispersion on Information Gathering



Note. This figure offers a 3D perspective on Figure E6, showing that information gathering decreases with stakes for low cost but increases with stakes for high cost. This includes the low-stakes condition only from Experiment 1, for comparison with Experiment 2. See the online article for the color version of this figure.

Figure E6

Interaction Between Cost and Dispersion on Information Gathering



Note. (A) The model predicts a two-way interaction whereby information gathering decreases with dispersion at low cost but increases with dispersion at high cost. The same interaction is observed between, but not within, groups in Experiment 2. The inflection point of the interaction appears to be the absolute level of information gathering, centered around six clicks (corresponding to TTB-like attribute-based processing; dashed line). (B) The same predictions are validated in Experiment 1, with information gathering converging toward six clicks as dispersion increases, regardless of the cost of clicking. Experiment 1 data are for the low-stakes condition only, to facilitate comparison with Experiment 2. Error bars show the 95% confidence interval across participants. TTB = take-the-best; exp. = experimental. See the online article for the color version of this figure.

(*Appendices continue*)

Appendix F

Decision Quality and Sources of Underperformance

This appendix provides additional details to accompany the sections on Decision Quality and Sources of Underperformance for each experiment in the main text.

Experiment 1

Performance

Here, we provide additional statistical results and figures that show decision quality when excluding low-effort participants and across all 50 conditions. Because group underperformance may be driven by low-effort participants who simply do not perform the task, we measured decision quality after excluding participants who gambled randomly on more than half of all trials ($n = 394$ or 16.6% of participants). As illustrated in Figure F1, the average decision quality of the remaining participants was 0.643, suggesting that the relatively low performance could not be fully (or even mostly) explained by low-effort participants (compare to Figure 6). The model's decision quality on the trials of attentive participants (0.907) was very similar to its decision quality on the trials of all participants. This suggests that at least 26% of the gap between attentive participants' performance and the performance of the unboundedly optimal decision strategy is due to people's sensitivity to click costs (which we use as a proxy for limited cognitive resources and opportunity costs), whereas at most 74% are due to people's deviations from resource-rational decision making. These numbers are only a lower/upper bound because future improvements to our resource-rational model, such as taking into account that people's utility function may be nonlinear (Kahneman & Tversky, 1979), or the experimental paradigm (see Experiment 2) could further increase the proportion of people's underperformance that the model can explain.

For attentive participants, we observed a similar pattern of changes across stakes, dispersion, and cost, as we did for all participants: $B = 0.029$, $p = .0096$, $B = 0.05$, $p < .001$, and $B = -0.046$, $p < .001$, respectively.

Sources of Underperformance

Here, we present the same results presented in the section on Sources of Underperformance for Experiment 1, but excluding low-effort participants who gambled randomly on more than half of all trials.

Figure F1 shows that, when excluding low-effort participants, the remaining participants achieved 70.9% of the gross performance of the model, which corresponds to 24.1 fewer points per trial on average. To account for this discrepancy, we measured four sources of underperformance: implicit costs of information gathering, imperfect use of the gathered information, imperfect strategy selection, and imperfect strategy execution. As shown in Figure F4, participants achieved 70.6% (95% CI [68.3, 71.2]) of the net performance of the model, with the four sources of underperformance, respectively, accounting for 4.2%, 4.7%, 14.6%, and 5.9% of the remaining 29.4% performance gap.

We estimated the implicit cost of information gathering as before, to control for the amount of information collected by participants and the model, resulting in an implicit cost of clicking of 1.5 points per click when excluding participants, and a 4.2% reduction in model performance (Figure F4). Notably, the contribution from random gambling drops considerably to 2.5%, and the overall contribution of imperfect strategy selection drops to 14.6% when excluding participants, which is still higher than the contributions of the other three sources of underperformance, but not as high as without participant exclusions (compare to Figure 7). Figures F4 and F5 show the same analyses presented in Figures 7 and 8, respectively, when excluding participants.

Experiment 2

Performance

Participants in the experimental group on average showed *worse* imperfect use of information than participants in the control group (see Figure 14), which indicates that not all participants were performing the task, since they were given the exact values to make perfect use of information (i.e., the subjective expected value of each gamble, see Figure 9). This is actually not surprising, considering that, in Experiment 1, 16.6% of participants gambled randomly (without gathering information) on more than half of all trials; in Experiment 2, participants were forced to wait for 20 s before gambling and therefore did not have the option to immediately gamble randomly, as they would in Experiment 1 or in the control group. To remove poor performers from both groups, we first computed the fraction of participants in the control group who gambled randomly on more than half of all trials (27.2%), to find comparable levels of poor performers across experiments (since the tasks were identical in Experiment 1 and the control group in Experiment 2). We then used this value to remove the bottom 27.2% of performers from each group in Experiment 2, but since the fraction of trials with random gambling is no longer a valid metric, we simply excluded participants based on their net performance as a fraction of the model's net performance. Figure F6 shows that attentive participants in the experimental group outperformed attentive participants in the control group in every condition, LD-LC: $t(55) = 2.36$, $p = .022$, $d = 0.63$; LD-HC: $t(80) = 4.02$, $p = .001$, $d = 0.90$; HD-LC: $t(78) = 2.26$, $p = .027$, $d = 0.51$; HD-HC: $t(73) = 2.30$, $p = .024$, $d = 0.53$.

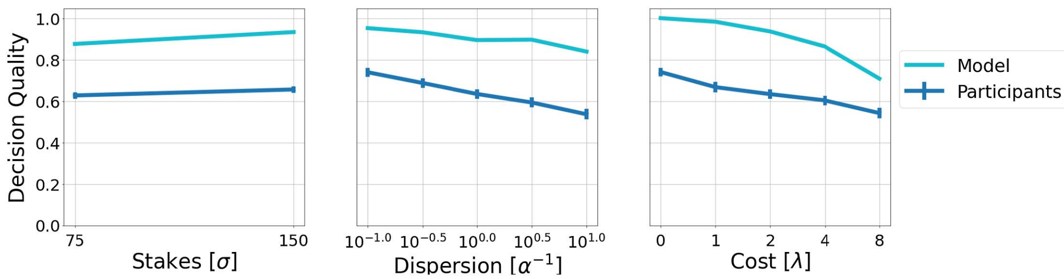
Sources of Underperformance

We computed the same four sources of underperformance after excluding low-effort participants from both groups. These results are shown in Figures F7 and F8. The fit implicit cost of gathering information was 0.2 and 1.9 points per click for the experimental group and the control group, respectively.

(*Appendices continue*)

Figure F1

Performance When Excluding Participants Who Gamble Randomly on More Than Half of All Trials From Experiment 1



Note. Performance was measured as the relative reward earned on each trial (the fraction of the highest possible reward with perfect information, omitting click costs). Error bars show the 95% confidence interval across participants. See the online article for the color version of this figure.

Table F1

Statistical Results Accompanying Figure 6 From Experiment 1

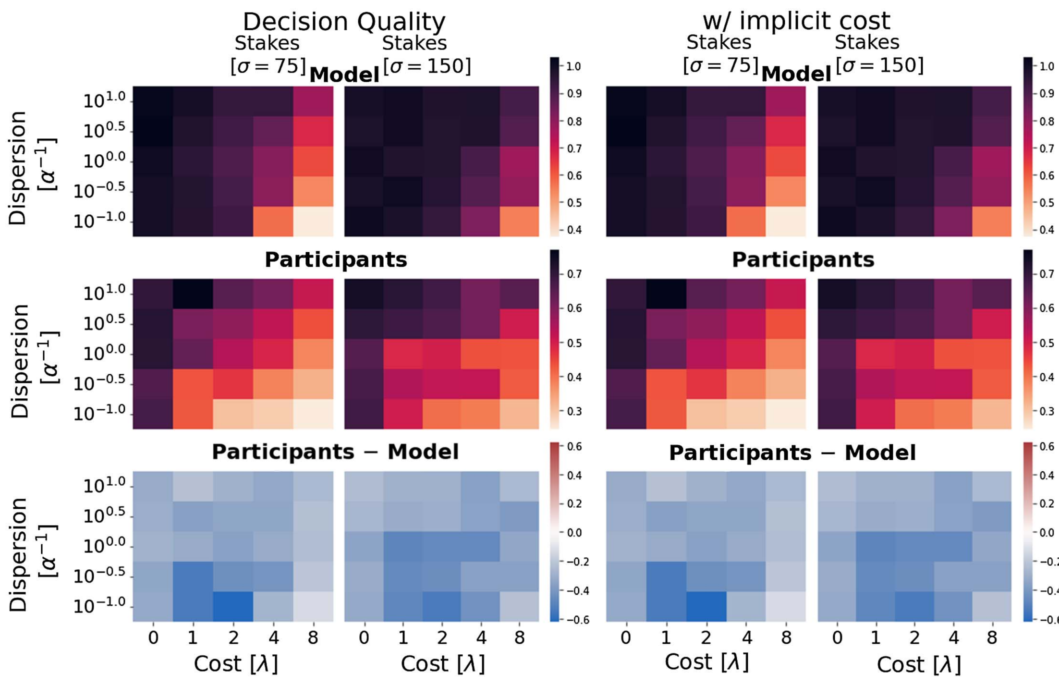
Behavioral feature	Independent variable	Main effect	Significant post hoc comparison	Effect size (Cohen's d)
Relative performance	Stakes	$t(2366) = -2.92$, $p = .0036$	n/a	−0.12
Relative performance	Dispersion	$F(4, 2363) = 42.76$, $p < .001$	All pairs except $10^{-0.5}$ and $10^{0.0}$	−0.22, −0.11, −0.24, −0.19
Relative performance	Cost	$F(4, 2363) = 50.48$, $p < .001$	All pairs except 1 and 2, 2 and 4	0.36, 0.16, 0.13, 0.2
Relative performance (with exclusions)	Stakes	$t(1972) = -2.59$, $p = .0096$	n/a	−0.12
Relative performance (with exclusions)	Dispersion	$F(4, 1969) = 43.29$, $p < .001$	All pairs except $10^{-0.5}$ and $10^{0.0}$	−0.23, −0.17, −0.22, −0.22
Relative performance (with exclusions)	Cost	$F(4, 1969) = 38.00$, $p < .001$	All pairs except 1 and 2, 2 and 4	0.3, 0.14, 0.13, 0.26

Note. Summary of statistical results corresponding to the analyses shown in Figure 6 from Experiment 1. A two-sample t test was used to test the main effect of stakes. Analyses of variance were used to assess the main effects of dispersion and cost. When applicable, post hoc pairwise comparisons were conducted between all 10 pairs of levels of each independent variable using two-sample t tests with Tukey's correction for multiple comparisons. These tests were not applicable (n/a) when the independent variable had only two levels or its main effect was not significant. The effect sizes for these comparisons were calculated using Cohen's d and are presented in ascending order of the corresponding levels of the independent variable (reporting adjacent pairs only).

(*Appendices continue*)

Figure F2

Decision Quality (Left Panels) and Decision Quality for the Model With an Implicit Cost of Clicking (Right Panels) Shown Across All 50 Conditions of Experiment 1, for the Model (Top Row), Human Participants (Middle Row), and the Difference Between the Model and Participants (Bottom Row)

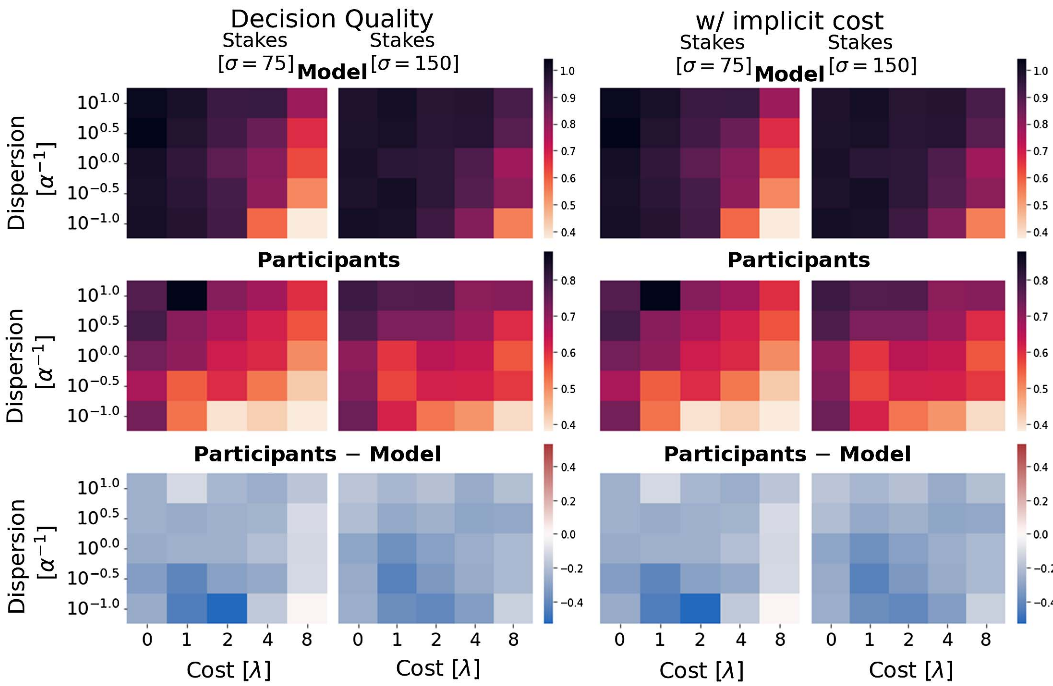


Note. The 50 conditions vary three parameters for a 2 × 5 × 5 across-participant design: σ (reward stakes), α^{-1} (uniformity of outcome probabilities), and λ (cost per click). The results here accompany the behavioral results shown in Figure 6. Within each parameter value in Figure F8, results are averaged across all values of other parameters, whereas in this figure, the full results for each of the 50 conditions are shown. See the online article for the color version of this figure.

(*Appendices continue*)

Figure F3

Same as Figure F2, but Excluding Participants Who Gambled Randomly on More Than Half of All Trials (n = 394 of 2,368 Participants Total) From Experiment 1

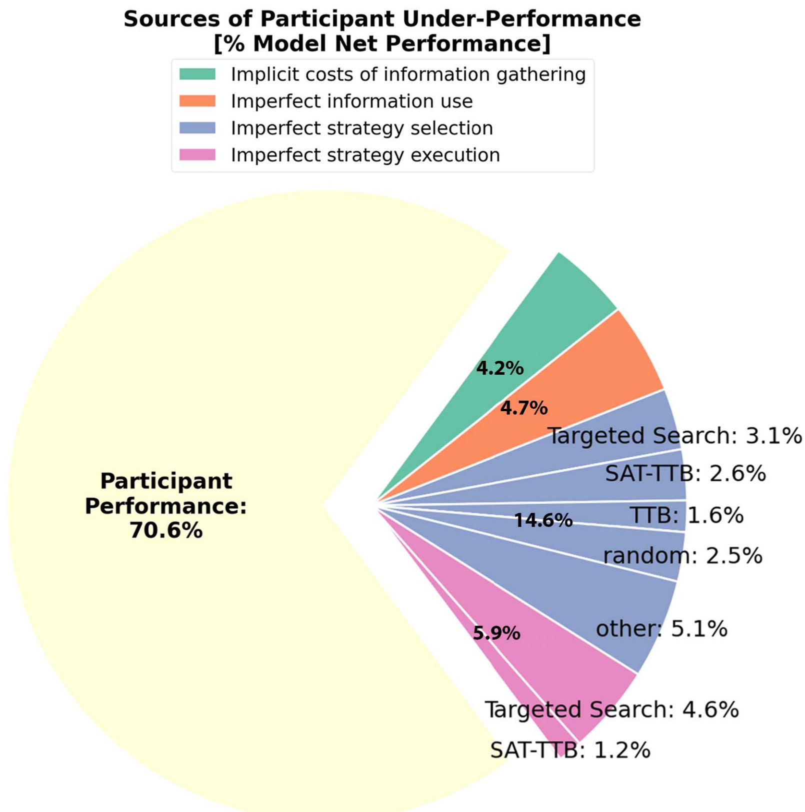


Note. See the online article for the color version of this figure.

(*Appendices continue*)

Figure F4

Sources of Underperformance When Excluding Low-Effort Participants From Experiment 1

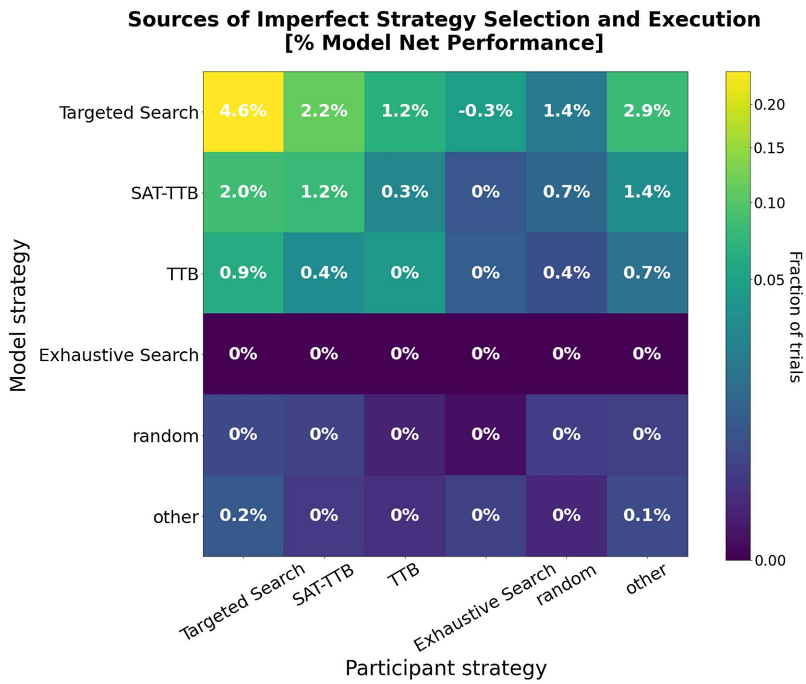


Note. Participants' net performance was 70.6% (95% CI [68.3, 71.2]) that of the model, with four distinct sources of the remaining 29.4% gap depicted here. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

(*Appendices continue*)

Figure F5

Sources of Imperfect Strategy Selection and Execution When Excluding Low-Effort Participants From Experiment 1

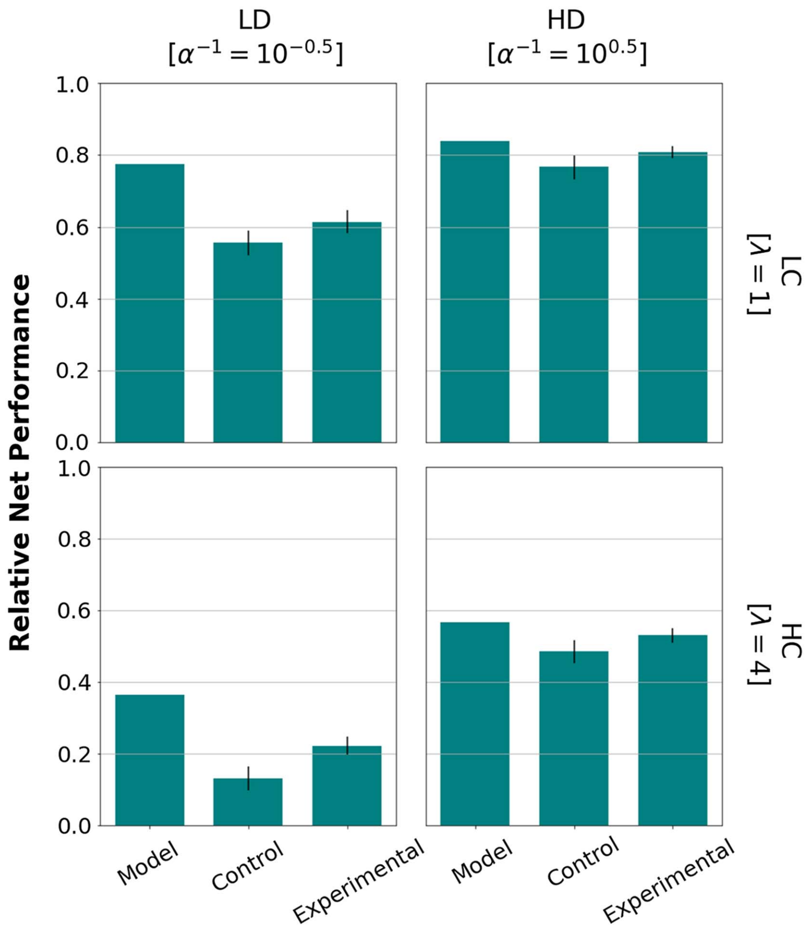


Note. Each cell states participants' average reduction of net performance from a trial-wise comparison of model–participant strategy selection. Off-diagonal cells correspond to imperfect strategy selection, while on-diagonal values correspond to imperfect strategy execution. Colors correspond to the number of trial-wise model–participant strategy pairs. See Figure 8 for details. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

(*Appendices continue*)

Figure F6

Performance Across Conditions for Each Group in Experiment 2 When Excluding Low-Effort Participants

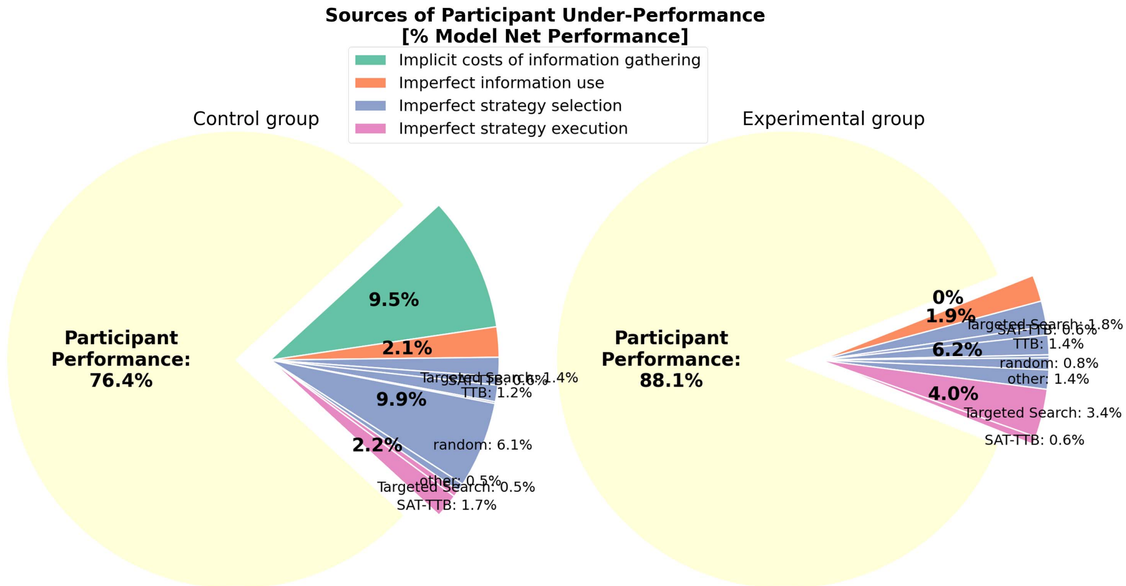


Note. Net relative performance, which accounts for the cost of gathering information, shows that participants in the experimental condition performed significantly better than participants in the control group in every condition. Error bars show 95% confidence interval across participants. LD = low dispersion; HD = high dispersion; LC = low cost; HC = high cost. See the online article for the color version of this figure.

(*Appendices continue*)

Figure F7

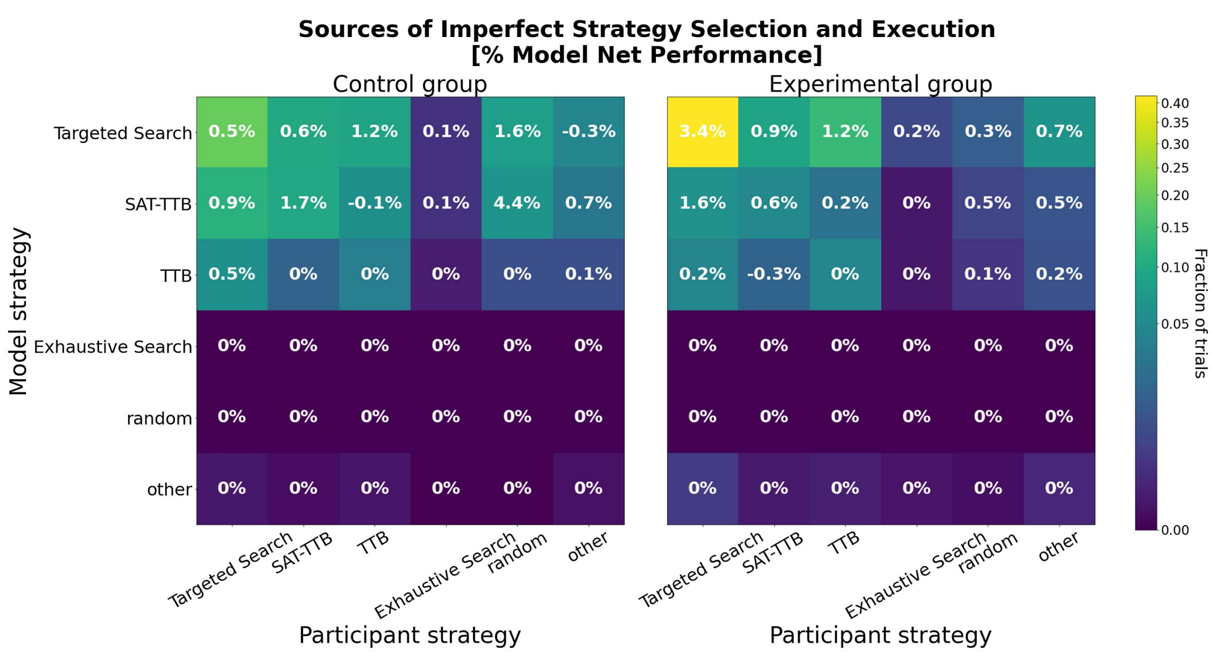
Same Results as Figure 14 From Experiment 2, but Excluding Low-Effort Participants



Note. Overall performance was 76.4% (95% CI [68.6, 80.4]) and 88.1% (95% CI [82.1, 91.8]) for the control group and experimental group, respectively. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

Figure F8

Same Results as Figure 15 From Experiment 2, but Excluding Participants Who Did Not Perform the Task Correctly



Note. TTB = take-the-best; SAT-TTB = satisficing-TTB. See the online article for the color version of this figure.

Received September 16, 2022
Revision received September 25, 2023
Accepted October 7, 2023 ■