Identifying Resource-Rational Heuristics for Risky Choice

Paul M. Krueger[1,a], Frederick Callaway[2,a], Sayan Gul[3],

Thomas L. Griffiths[1,2,b], and Falk Lieder[4,b]

[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

Author Note

Correspondence concerning this article should be addressed to Paul M. Krueger, Department of Computer Science, Princeton University, 35 Olden St., Princeton, NJ 08540. E-mail: paul.m.krueger@gmail.com. [a]PMK and FC contributed equally to this work. [b]TLG and FL share joint senior authorship on this article. 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, GK. This material has been substantially revised and expanded for the present article. This work was supported by grant number MURI N00014-13-1-0341 from the Office of Naval Research, grant number 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 declare no conflict of interest. All code and data used to run the experiments and produce the results presented in this paper are available at https://github.com/fredcallaway/rational-heuristics-risky-choice/.

Abstract

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

Identifying Resource-Rational Heuristics for Risky Choice

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 & Murawski, 2017; Bossaerts et al., 2019; Gershman et al., 2015; Kwisthout et al., 2011; Nowozin, 2014; Papadimitriou & Tsitsiklis, 1986). According to classic economic theory, people should choose their actions so as to maximize the expected value of the consequences (Morgenstern & Von Neumann, 1953; Savage, 1951), but computing those expected values 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 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., 2015; Griffiths et al., 2012; 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 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 strategy. To operationalize the cost of gathering information, 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.

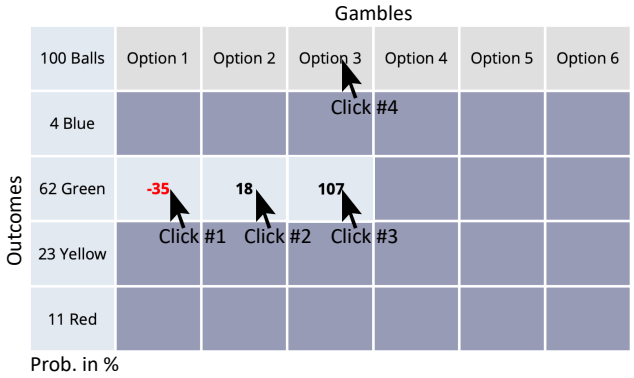

Figure 1

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

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 (Gigerenzer & Goldstein, 1999) and Weighted-Additive (Dawes & Corrigan, 1974; Payne et al., 1988) heuristics as resource-rational strategies in specific situations, validating the approach. In addition, our method discovered novel heuristics 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 Weighted Additive model, in which payoffs are weighted by their probabilities.[1] Another widely-recognized heuristic is the lexicographic rule (Svenson, 1979; Tversky, 1969) or "Take-The-Best" (Gigerenzer & Goldstein, 1999), which focuses on a single

[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).

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, et al., 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 Mousleab task, may incur only a minimal reduction in accuracy. For example, when one attribute is much more likely than all the others, Take-The-Best performs nearly as well as the much more costly Weighted Additive 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 vs. 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) assume 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 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 externalize the cognitive cost of gathering information directly, which allows us to compute precisely the optimal trade-off between effort (operationalized as click costs) and decision accuracy. This provides 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, 2002; Martignon et al., 1999; 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 account for the effectiveness of heuristics in terms of the structure of the decision environment (Baucells et al., 2008; Bhatia & Stewart, 2018; Dieckmann & Rieskamp, 2007; Katsikopoulos, 2011; Katsikopoulos & Martignon, 2006; Martignon & Hoffrage, 2002; Martignon et al., 1999; 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), the effectiveness of reducing model parameters to balance the bias-variance trade-off (Gigerenzer & Brighton, 2009; Holte, 1993) or when observations are limited or noisy (Hogarth & Karelaia, 2005, 2006, 2007; Şimşek & Buckmann, 2015), and Bayesian inference with strong priors (Parpart et al., 2018). More recently, a resource-rational analysis of cognition has been applied to view heuristics as making 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 elementary information processes (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 paradigm relies on the MDP framework as a formal representation of the external environment and 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-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, et al., 2021) as well as planning (Callaway, Lieder, et al., 2018; 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 elementary information processes (detailed above; Bettman et al., 1990; Johnson

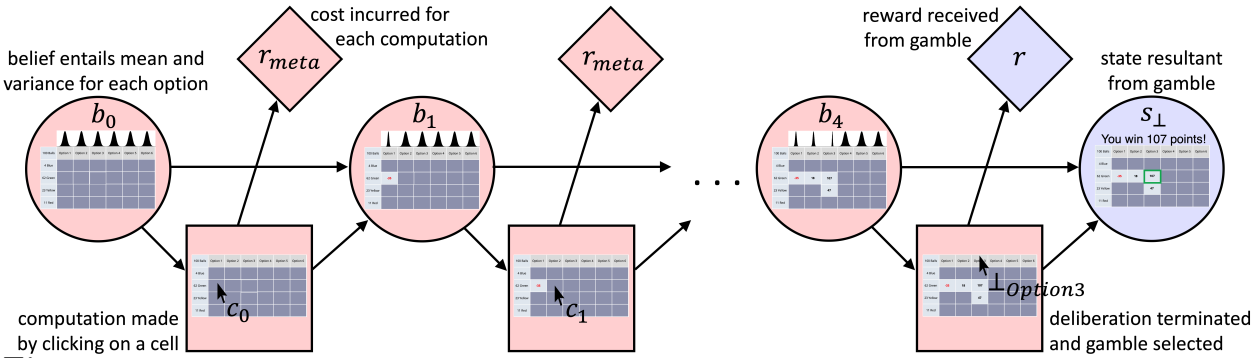

Figure 2

Schematic illustration of the meta-level Markov Decision Process framework applied to the Mouselab task. 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).

& 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-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, Gul, 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. Finally, 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 posterior over the value of each gamble given the currently revealed values, and computations correspond to revealing a cell and updating the posterior accordingly. Solving this meta-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 were to occur. Critically, all the payoffs are initially occluded, and the participant has to click on a cell to reveal its entry. The probabilities of the different outcomes are known to the participant. Each click comes at a cost, and participants are free to inspect as many or as few cells as they would like.

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: 1) the "stakes" of the decision (the variance of possible payoffs), 2) the "dispersion" of the outcome distribution (lower values resulting in more similar probabilities for each outcome), and 3) the "cost" of computation (the number of points subtracted for each click). We considered two levels of stakes and five levels for dispersion and for 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 Markov Decision Processes (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 so as 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 (that is, 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 $(b_{t,1}, \cdots, b_{t,n})$ 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 according to Bayesian inference. Since the entries of the payoff matrix are drawn from the Gaussian distribution $\mathcal{N}(\bar{v}, \sigma_v^2)$, the resulting posterior distributions are also Gaussian. Hence, the decision-maker's belief about the expected payoff of the g^{th} 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.

The meta-level transition function $T_{\text{meta}}(b_t, c_{o,g}, b_{t+1})$ encodes the probability distribution on what the updated means and variances will be given the observation of a payoff value $v_{o,g}$ sampled from $\mathcal{N}(\bar{v}, \sigma_v^2)$, and is determined using Bayesian inference integrating over the distribution of possible observed payoff values. 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 $r_{\text{meta}}(b_t, \perp) = \max_g b_t^{(\mu)}(g)$, 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 $h^\star$ 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 an well-informed belief state minus the cost of attaining it, that is

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

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

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 Bayesian meta-level policy search algorithm (Callaway, Gul, et al., 2018). In Appendix A we provide details of how this algorithm can be applied to find near-optimal strategies in this model.

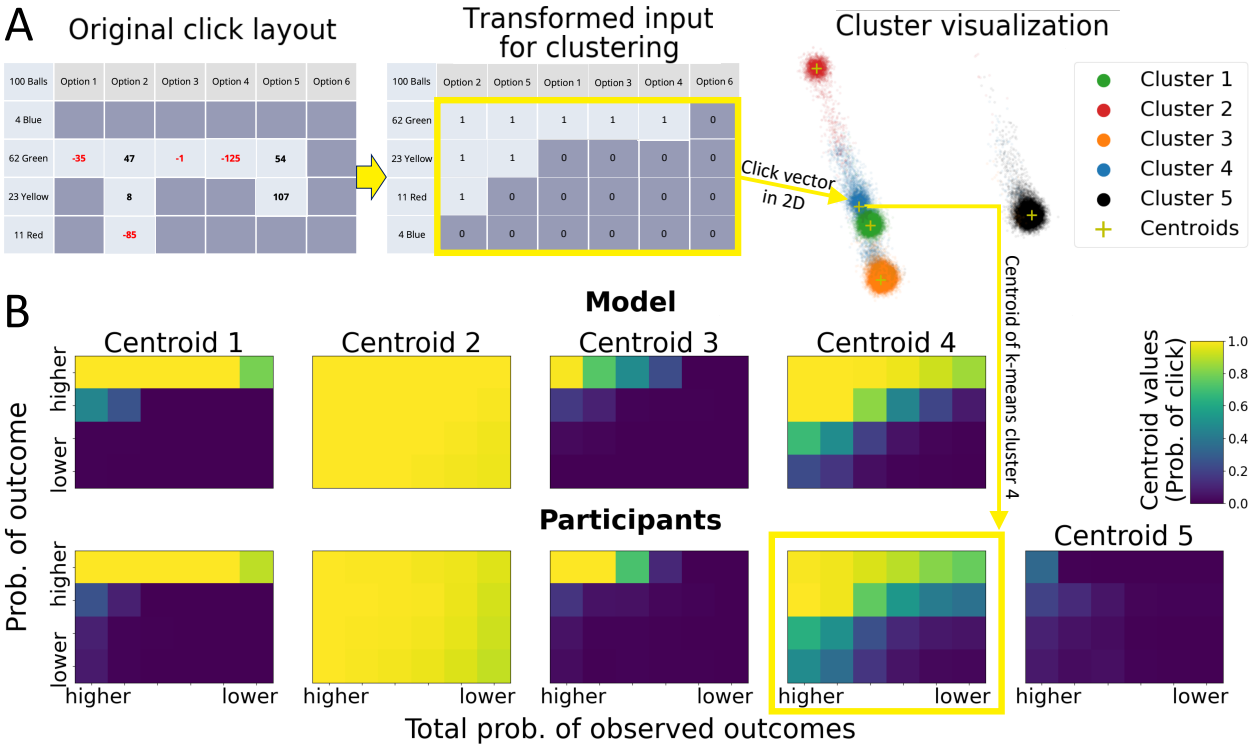

Figure 3

Identification of heuristics. ***(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 Fisher's Linear Discriminant Analysis) 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 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 SAT-TTB+. A fifth cluster corresponding to gambling randomly (without gathering information) was also revealed in the human data.*

Identification of resource-rational heuristics

As discussed above, previous work has identified a set of well-known heuristics that people use in multi-attribute risky choice. For example, Take-the-Best (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 is good enough (Simon, 1956); it is sometimes referred to as a conjuctive rule. These heuristics both ignore information about some alternatives or attributes. In contrast, a less frugal strategy, Weighted Additive (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 heuristics encompassed by all the wide range of decision environments we considered.

We found the best strategy for each of 1000 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 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. Finally, 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

[2] If there is a tie, then TTB considers the second most predictive attribute (and so on) but this scenario virtually never occurs in our paradigm because there are about 1000 possible payoffs.

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 1000 problems, with a Euclidean distance metric (Elkan, 2003). In this and all subsequent analyses, the distribution of the 1000 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. Fisher's Linear Discriminant Analysis (LDA) was used to project the 24-dimensional click sequence vectors onto a 2-dimensional space (Fisher, 1936). 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), which 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 the prototypes for the resource-rational model in centroids 1 and 2 revealed that our method rediscovered the TTB heuristic (Gigerenzer & Goldstein, 1999) and the WADD strategy, respectively. This indicates that these heuristics strike a near-optimal trade-off between decision quality and cognitive cost, at least in some situations. TTB corresponds to inspecting only the most probable attribute for each alternative gamble. The WADD strategy clicks practically everywhere, hence the nearly all-yellow color. Rediscovering these classic heuristics provides support for the validity of our approach.

Centroids 3 and 4 correspond to the prototypes of two newly discovered strategies. The first, which we call SAT-TTB, combines elements of TTB and Satisficing (see Figure 1), and may be likened to a single-dimension conjunctive rule. 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. The second newly discovered heuristic, SAT-TTB+, starts by inspecting some or all of the payoffs for the most probable outcome (as in SAT-TTB), and 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 Figure 2 and in Figure 3A). This strategy is similar to a lexicographic semi-order model (e.g., Birnbaum & Gutierrez, 2007; Manzini & Mariotti, 2012; Safarzadeh & Rasti-Barzoki, 2018; Tversky, 1969). The two newly discovered heuristics do not correspond to any heuristics previously observed in Mouselab. Yet, as described below, we found that people frequently use these heuristics across the wide range of environments in which they are adaptive.

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

Methods

Participants

We recruited 2,368 participants on Amazon Mechanical Turk (1,115 females, mean age 37.6 years, standard deviation 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 (standard deviation 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 fifty 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 1) 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 $\sigma \in \{75, 150\}$), 2) 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 3) 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 re-read the instructions and retake the quiz. If they failed the quiz three times, they were not allowed to participate in the main task.

Transparency and openness

Our results did not exclude any participants (except where 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 paper 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 heuristic use to the statistics of the environment, and then compared this to how people's heuristics 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 mode. Data from each trial was transformed in exactly 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 perfectly those produced by the model, manifesting the classic strategies TTB and WADD as well as the newly discovered SAT-TTB and SAT-TTB+. While the resource-rational model never gambles randomly, participants do occasionally gamble without gathering any information; this is captured in centroid 5.

Based on the clustering solution, we defined 5 distinct strategies to be considered in subsequent analyses as follows: 1) SAT-TTB+ 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; 2) 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; 3) TTB was operationalized as selecting all 6 cells from the most probable row, and nothing else; 4) WADD was defined as selecting all 24 cells; 5) 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.

Comparison of strategies across environments

The clustering results indicate that people use the same types of heuristics as the resource-rational model. To determine whether people deploy these heuristics 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 SAT-TTB+, 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 ($B = -2.3, p < 0.001$) and a significant positive effect on the frequency of SAT-TTB+ ($B = -1.3, p < 0.001$; 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

[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.

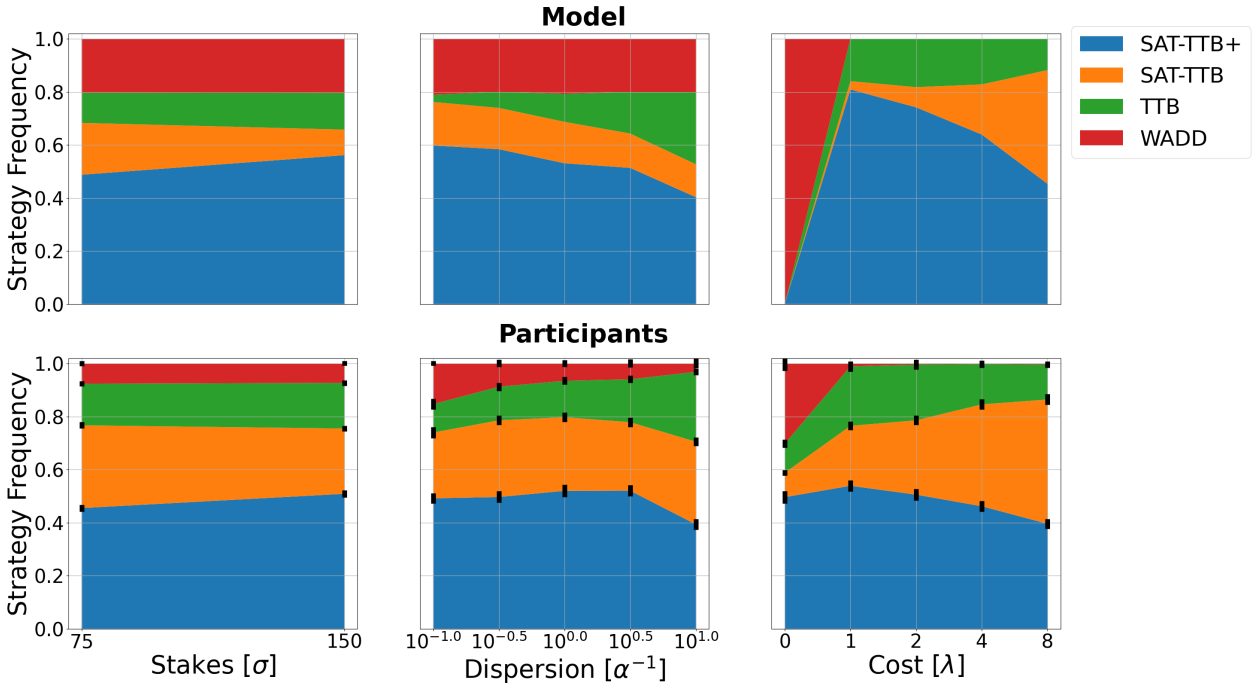

Figure 4

Use of Weighted Additive (WADD), Take-The-Best (TTB), and variations of satisficing-TTB (SAT-TTB and SAT-TTB+) 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% CI across participants

single outcome when only one is likely to occur. Our participants confirmed this prediction ($B = -5.5, p < 0.001$; middle column of Figure 4). However, while the resource-rational model most-often uses SAT-TTB+ in low-dispersion environments, participants often resorted to choosing randomly instead ($B = -1.4, p < 0.001$).

When there is no cost for gathering information, the model always uses WADD since the value of information is always positive. Although participants also limited their use of WADD to this case, they were more likely to use SAT-TTB+. 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 SAT-TTB+ ($B = -0.8, p < 0.001$) and TTB ($B = -3.9, p < 0.001$), while increasing use of the most frugal strategy, SAT-TTB ($B = -3.3, p < 0.001$). Figures C1 and C2 compare strategy frequencies in each

of the 50 conditions, showing broad correspondence between the resource-rational model and participants.

Table C1 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 5A, 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

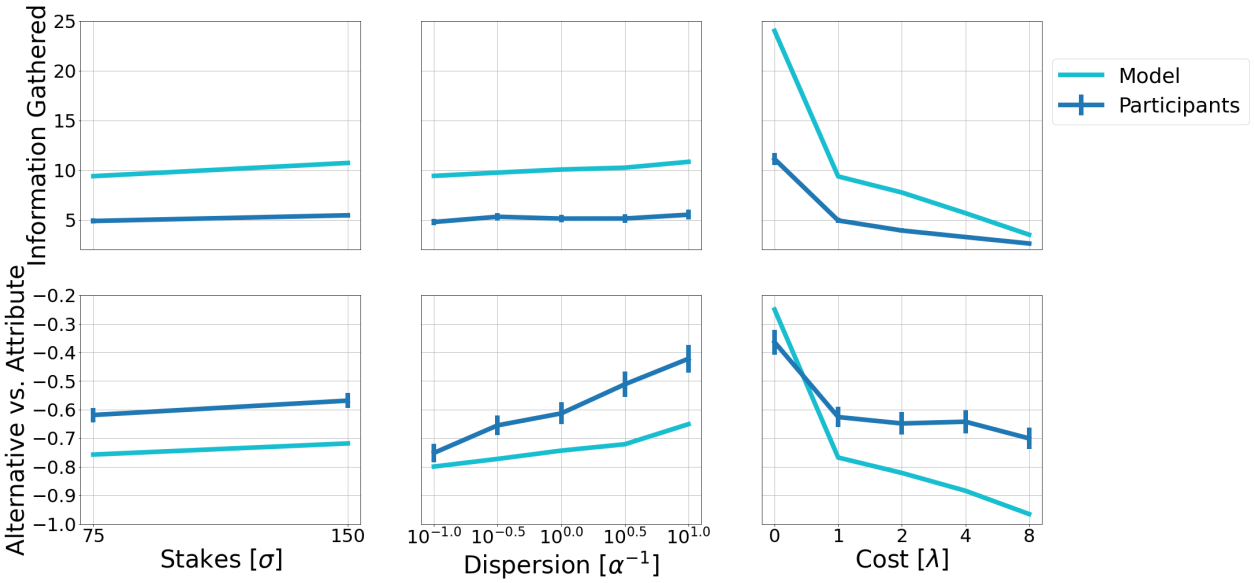

Figure 5

Behavioral correspondence between participants and the resource-rational model in Experiment 1. ***(A)*** *The average number of values revealed by participants and the model as a function of each environment parameter.* ***(B)*** *The same, but for a measure of alternative- vs. attribute-based processing (negative indicates attribute-based). Error-bars show the 95% CI across participants.*

regression with random intercepts for participants revealed that the stakes significantly predicted information gathered: $B = 0.57, p = 0.009$). 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 ($B = -0.13, p = 0.097$). Finally, people reduced information gathering as it became more costly to do so ($B = -1.9, p < 0.001$). 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 vs. 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 5B 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, 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 ($B = 0.052, p = 0.016$); they used more attribute-based processing as dispersion increased ($B = -0.084, p < 0.001$) and as cost increased ($B = -0.073, p < 0.001$). A comparison of information gathering and alternative- vs. attribute-based processing for the model and participants across each of the fifty decision environments shown in Figure D1, showing an overall qualitative correspondence.

Two additional informative behavioral markers are the variance in the amount of information gathered across outcomes and across 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 "non-compensatory" 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 SAT-TTB+) or stop searching once a high-value gamble is found (as in SAT-TTB). Figure D2 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; $B = -0.01, p < 0.001$) and alternatives (alternative variance; $B = -0.004, p = 0.0016$), 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 ($B = 0.0091, p < 0.001$) and more evenly across the alternatives ($B = -0.0026, p < 0.001$). Finally, increasing the cost of information made people more discerning in how much attention they paid to different attributes ($B = 0.017, p < 0.001$) and different alternatives ($B = 0.009, p < 0.001$). 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 D3 shows the qualitative correspondence between the model and participants for these two measures across all fifty 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 therefor multiple samples within a single alternative become less useful, consistent with more frequent use of TTB and less frequent use of SAT-TTB+ 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 D1 presents results of additional statistical tests for the results reported in this section.

Performance

In addition to providing a framework for discovering heuristics, 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 vs. genuinely irrational use of one's cognitive resources. We measured people's relative performance by the fraction of the highest expected reward attainable with perfect information, omitting the cost of information gathering. This relative measure allows as to compare resource-rational performance of the model to unboundedly optimal performance with perfect information (i.e., maximum expected value). Omitting the cost of information gathering is useful for this comparison, and for comparing gross performance across cost conditions.

As illustrated in Figure 6, our resource-rational model performs relatively close to the unboundedly optimal standard of 1.0, falling shorter when less information is gathered. Concretely, the average relative performance was 0.895. Furthermore, our model accurately predicted that participants' relative performance increases with the stakes ($B = 0.039, p = 0.0035$), decreases with the dispersion of the outcome probabilities ($B = 0.059, p < 0.001$), and decreases with the cost of gathering and processing information ($B = -0.063, p < 0.001$). These results are shown in Figure 6.

It is apparent that participants under-perform compared to the model. The average relative performance of all participants was only 0.542. Thus, 23% of the gap between all participants' performance and the performance 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.

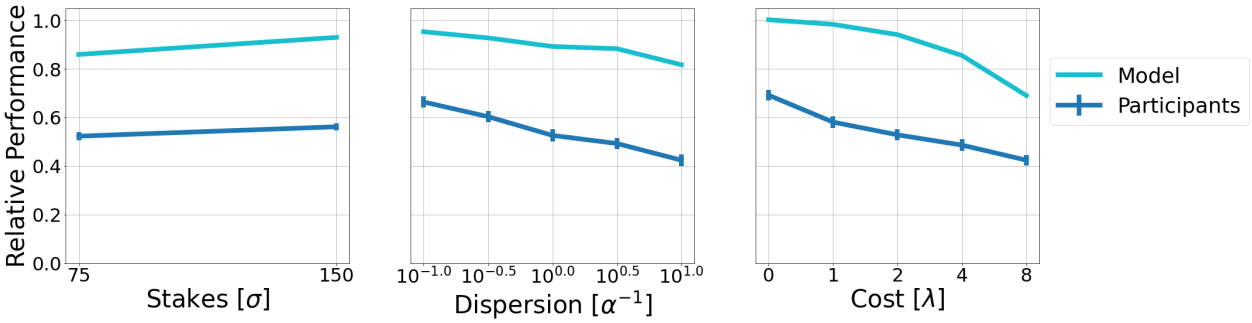

Figure 6

Participants show a qualitative correspondence to the model in performance across conditions in Experiment 1. 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% CI across participants.

Table E1 summarizes the main effects, corrections for multiple comparisons, and effect sizes for measuring relative performance across conditions. Figure E2 shows a qualitative correspondence between participants' and the resource-rational model's relative performance across all 50 environmental conditions; see Appendix E for detailed results when excluding low-effort participants who gambled randomly on more than half of all trials (16.6% of participants).

Sources of under-performance

Our resource-rational model allowed us to investigate how close human performance comes to the upper bound established by our resource-rational model. As shown in Figure 6, people performed systematically worse than the resource-rational model across all environments. Participants' average relative performance was 60.6% that of the model, as shown in Figure 6. Measured in raw points, participants achieved an average of 31.8 fewer points per trial than the resource-rational model. What explains this sizable gap? There are at least four possible reasons why people might be suboptimal: implicit costs of information gathering, imperfect use of the gathered information, imperfect strategy selection, and imperfect strategy execution. As detailed below and illustrated in Figure 7, participants achieved 61.4% (95% CI [58.8, 61.9]) of the net performance of the model,

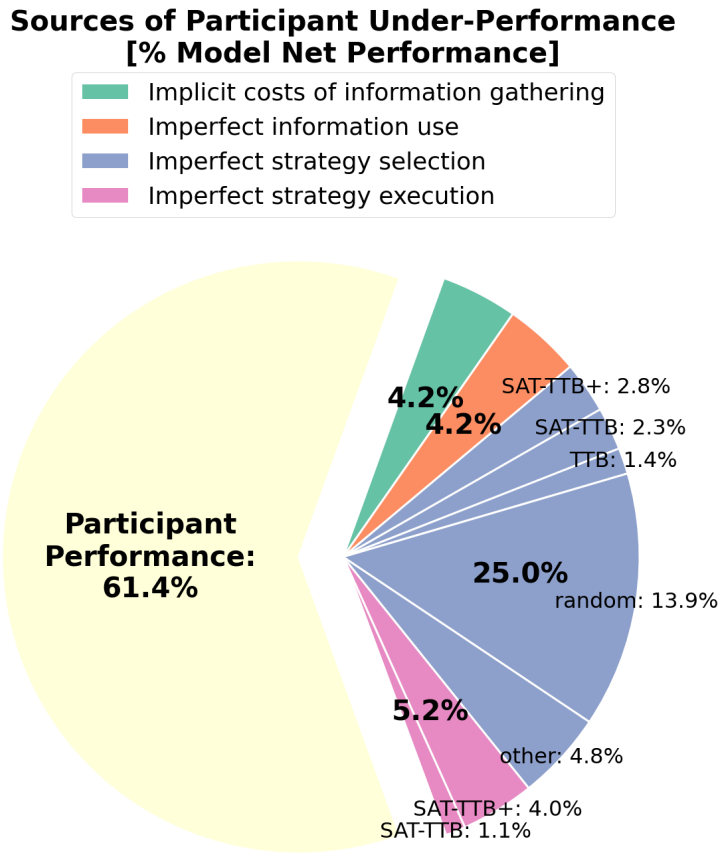

Figure 7

Sources of under-performance in Experiment 1. 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.*

with each of these four sources respectively accounting for 4.2%, 4.2%, 25.0%, and 5.2% of the remaining 38.6% participant under-performance (measured as a percentage of the model's net performance, as detailed below. See Appendix E for detailed results when excluding low-effort participants).

We now describe these four sources of under-performance, and assess the degree to which they contribute to people's under-performance in turn. The measure of relative performance plotted in Figure 6 omitted the costs of information gathering (to facilitate comparisons across conditions), and measured *relative* performance as a fraction of performance with *perfect information* (to assess the model's resource-rational performance

against an unboundedly optimal upper bound). Here (including Figure 7), we used *net* performance, defined as the payoff received minus the costs of information gathering, as a fraction of the *model's* performance. This measure of net performance allows us to compare people's resource rationality against the standard set by our normative model. This analysis revealed that, on average, the net performance of participants' decision strategies was 61.4% (95% CI [58.8, 61.9]) of the net performance of resource-rational decision-making. The four sources of under-performance collectively account for the remaining 38.6% gap in their performance.

First, participants may be influenced by costs not accounted for by our resource-rational model. This might be able to explain why participants collected less information than the resource-rational model. These costs might capture, for example, the cost of the time required to move a cursor and make clicks, as well as the anticipated cognitive costs associated with processing the revealed information (Payne et al., 1988). To assess the degree to which insufficient information gathering led to participants' suboptimal performance, we ran 1,000 simulations of the model on each of the exact same trials presented to human participants, and measured net performance. To control for the overall amount of information gathered between our method and participants, we fit an implicit cost of information gathering to match the average number of clicks made by participants using a grid search. We found that an implicit cost of 2.4 points per click led to the same amount of information gathering on average as participants. We then measured the net performance of the model with an implicit cost of clicking of 2.4, and found a 4.2% reduction from the model without an implicit cost (as depicted in Figure 7).

A second source of under-performance is the imperfect use of gathered information. That is, given the information revealed, participants may simply fail to select the gamble with the highest expected value. This shortcoming can be accounted for by the effort required to compute such values in this task. To measure the extent to which people are suboptimal because they make imperfect use of the collected information, we computed the

conditional expected values of all alternatives given the information revealed by the participant. We then compared participant net performance to what it would have been had they chosen the gamble with the highest information-contingent expected value. As with the previous source of under-performance, this difference in net performance was measured as a percentage of the resource-rational model's net performance.

A third possible source of under-performance is imperfect strategy selection. At an aggregate level, people use the same heuristics 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 heuristic. We compared the strategy selected by participants on each trial to that chosen by the model. To measure imperfect strategy selection, 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 under-performance (by conditioning on the amount of information gathered and the use of information).[4] As before, this reduction in net performance was quantified as a percentage of the net performance of the resource-rational model.

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 SAT-TTB+. Such imperfect strategy execution is the fourth potential source of under-performance. To calculate this, 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 under-performance, and again measuring it as a percentage of the model's net performance.

Together, these four factors cover all possible ways in which people might deviate from the normative standard of resource-rational decision-making.

---

[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.

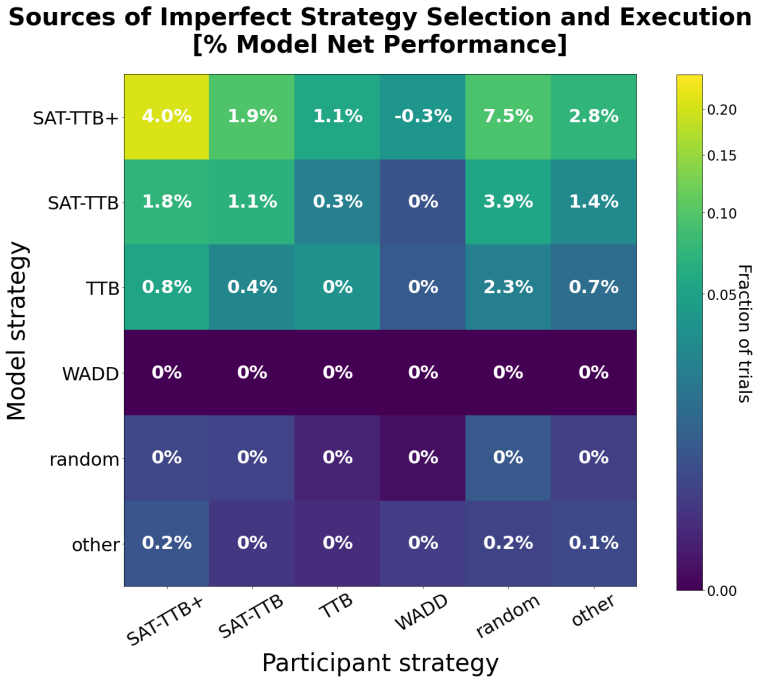

Figure 8

Sources of imperfect strategy selection and execution in Experiment 1. 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 SAT-TTB+ 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 SAT-TTB+ but the model chose SAT-TTB, with* 3,394 *such trials occurring (thus the teal color).*

Figure 7 shows the contribution of each of these four sources to under-performance. Implicit costs of gathering information account for 4.2% of participant under-performance. Participants failed to choose the gamble with the highest subjective expected value on 27.3% of all trials, losing 10.4 points on average on such trials. This imperfect use of information accounts for 4.2% of the participants' under-performance.

Imperfect strategy selection accounts for the majority of participant under-performance (25.0%), with random gambling accounting for most of it (13.9% of participant under-performance). Overall, these results suggests 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.

Finally, Figure 7 also shows that imperfect strategy execution contributes 5.2% to participants' under-performance. Errors in executing SAT-TTB+—the most complicated strategy—accounted for most of this source. Figure 8 displays the average reduction in performance based on a trial-wise comparison of participant and model strategies. Off-diagonal values correspond to imperfect strategy selection. For example, trials in which participants gamble randomly and the model chooses SAT-TTB+ account for 7.5% of under-performance, 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 SAT-TTB+, participants lost an average of 4.0% of the model's net performance. Colors depict the number of trials occurring for each participant-model strategy pair.

Consistent with the idea that people first choose a decision strategy and then execute it, we found that participants deliberated longer before the first click (2.92 sec) than before subsequent clicks (0.81 sec, $t(2549) = 128.5, p < 0.001$). Deliberation time also predicted information gathering, such that longer deliberation was followed by more frugal strategies (0.62 fewer clicks for each second spent deliberating; $B = -0.62$, $t(38737) = -37.6$, $p < 0.001$).

Discussion

While our resource-rational model successfully predicted how participants adapt their decision heuristics 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 under-performance.

The first source of under-performance—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.

It is worth noting that the physical effort required in our task to move the cursor and click on cells was a useful experimental adaptation to objectify the cognitive cost of information gathering. The Mouselab task typically reveals information when the cursor simply hovers over a cell, occluding it once again when the cursor leaves the cell, which is thought to mimic the real-world process of gathering information through eye movements. While there is mixed evidence whether information gathering in Mouselab differs significantly from eye movements (Glöckner & Betsch, 2008; Lohse & Johnson, 1996; Reisen et al., 2008), these other forms of information gathering could also in principle be captured in a more complicated meta-MDP model. Future work may apply our resource-rational approach to more precisely identify the underlying cognitive processes involved in deriving heuristic decision-making.

Experiment 2: Reducing cognitive constraints

To evaluate the extent to which participants' under-performance was due to implicit costs and cognitive limitations not accounted for by our model, we ran a second experiment with an experimental condition designed to remove or reduce some of those costs and mitigate those limitations. In particular, participants were forced to spend a minimum of

20 seconds on each trial, and the subjective expected value of each gamble given the observed information was displayed for each gamble and updated whenever new information was revealed. These two manipulations were intended to reduce the opportunity cost of the time it takes to obtain and process additional information and the cognitive cost associated with estimating the expected value of each gamble, respectively. We predicted that these manipulations would bring people's decision strategies into closer alignment with the predictions of our resource-rational model by reducing unaccounted costs of information gathering and helping people use the acquired information as effectively as our model does.

Methods

Participants

We recruited 404 participants on Amazon Mechanical Turk (250 males, mean age 37.5 years, standard deviation 10.8 years), 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 (standard deviation 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 8 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

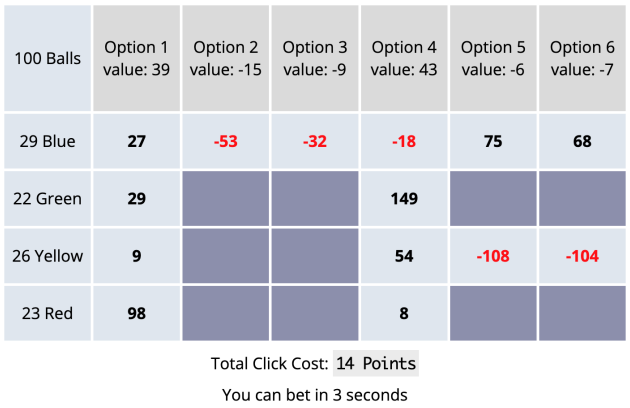

Figure 9

Screenshot from Experiment 2. To reduce implicit costs associated with information gathering and information use, participants in the experimental group were given a 20*-second time-minimum per trial, and a display of the subjective expected value of every gamble.).*

was updated according to Equation 3 and displayed atop that column. Furthermore, the experimental group was forced to spend a minimum of 20 seconds on each trial, and a countdown timer was displayed for the first 20 seconds of each trial. After the first 20 seconds, 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, $\$1.77 \pm \$0.97$) than participants in the control group (9.8 ± 5.2 min, $\$1.54 \pm \$0.95$).

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/.

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, $p < 0.001$; 4.5% vs. 27.2% of participants gambled randomly on more than half of all trials, $p < 0.001$). 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

[5] The clusters discovered for the model are not identical to those seen in Figure 3, corresponding to Experiment 1, because 1) 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 2) 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).

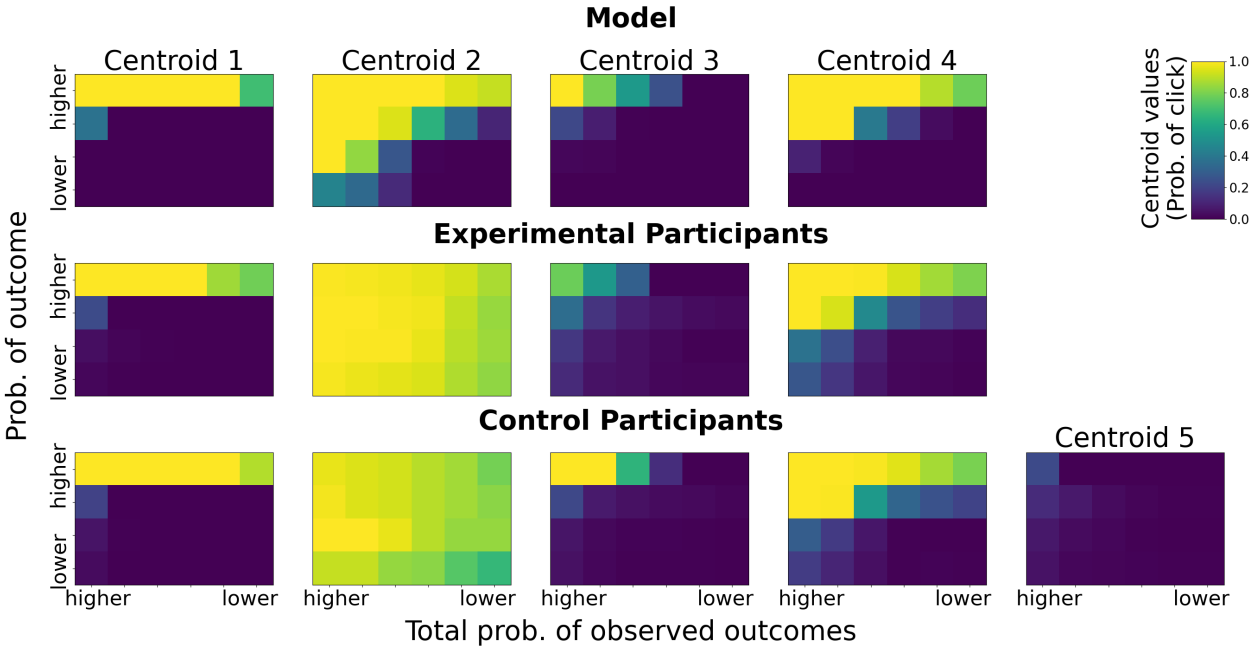

Figure 10

Experiment 2 k-means centroids. 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).*

an increase in the use of SAT-TTB+ (LD-LC: $\chi^2(1, 3960) = 32.9, p < 0.001, d = 0.25$; LD-HC: $\chi^2(1, 3960) = 49.9, p < 0.001, d = 0.32$; HD-LC: $\chi^2(1, 3960) = 32.9, p < 0.001, d = 0.25$; HD-HC: $\chi^2(1, 3960) = 32.9, p < 0.001, d = 0.25$). Conversely, participants used the more frugal SAT-TTB strategy less often in all environments except for the LD-HC environment (LD-LC: $\chi^2(1, 3960) = 12.4, p < 0.001, d = -0.16$; LD-HC: $\chi^2(1, 3960) = 0.1, p = 0.8, d = 0.01$; HD-LC: $\chi^2(1, 3960) = 12.4, p < 0.001, d = -0.16$; HD-HC: $\chi^2(1, 3960) = 12.4, p < 0.001, d = -0.16$). While these overall changes away from the frugal SAT-TTB heuristic toward the more costly SAT-TTB+ 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, WADD. While the model never uses WADD, participants in the experimental group used it more than those in the control

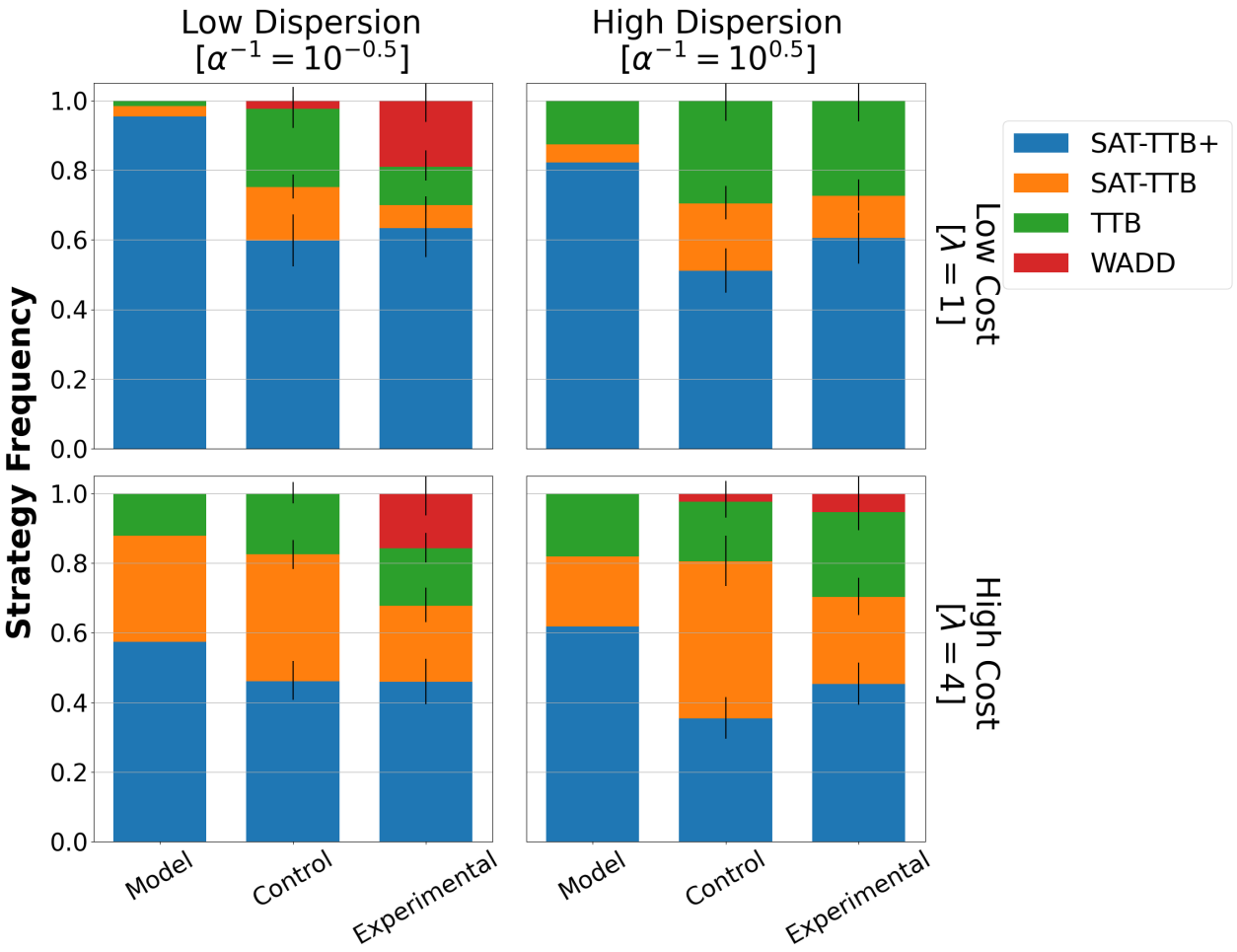

Figure 11

Reducing implicit costs increases the use of costly heuristics. Participants in the experimental group in Experiment 2 show a general increase in the use of SAT-TTB+, and even WADD, and a general decrease in the most frugal heuristic, SAT-TTB.

group in all environments except HD-LC (LD-LC: $\chi^2(1, 3960) = 114.9, p < 0.001, d = 0.53$; LD-HC: $\chi^2(1, 3960) = 124.6, p < 0.001, d = 0.69$; HD-LC: $\chi^2(1, 3960) = 114.9, p < 0.001, d = 0.53$; HD-HC: $\chi^2(1, 3960) = 114.9, p < 0.001, d = 0.53$). 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 C3).

Information gathering and choice behavior

The shift toward more costly heuristics in the experimental group is apparent in an overall increase in information gathering compared to the control group, shown in Figure 12. In each environment, participants in the experimental group gathered more information than those in the control group (two-sample t-tests; LD-LC: $t(102) = 4.88, p < 0.001, d = 0.96$; LD-HC: $t(100) = 4.71, p < 0.001, d = 0.93$; HD-LC: $t(100) = 3.23, p = 0.0017, d = 0.64$; HD-HC: $t(94) = 3.31, p = 0.0013, d = 0.68$). 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: $t(102) = -0.38, p = 0.7, d = -0.07$; LD-HC: $t(100) = 0.74, p = 0.46, d = 0.15$; HD-LC: $t(100) = -2.65, p = 0.0094, d = -0.52$; HD-HC: $t(94) = -0.45, p = 0.65, d = -0.09$).

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 D.

Performance

The model's relative performance was 0.886, that is, 88.6% of the gross performance of the unboundedly optimal strategy that always chooses the gamble with the highest expected value, with perfect information. Similar to Experiment 1, this percentage was only 47.9% in the control group. By contrast, for the experimental group, it was 67.8%. Excluding low-effort participants increased these percentages to 62.1% for the control group and 78.8% for the experimental group, with the model's relative performance equal to 87.9% for the same trials. This suggests that about 44% of the control group's total

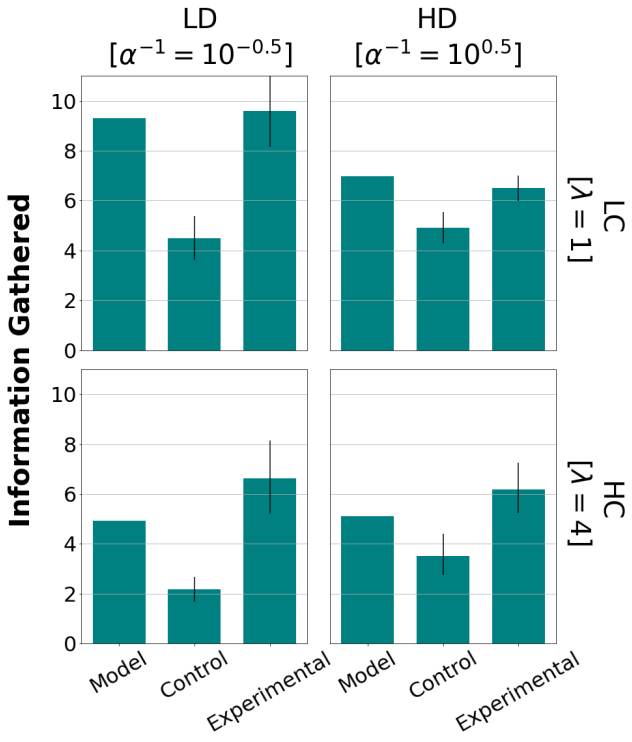

Figure 12

Information gathering for each group in Experiment 2. 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). (LD = Low Dispersion, HD = High Dispersion, LC = Low Cost, HC = High Cost; error-bars show 95% CI).

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. This suggests that people are more resource-rational than they appeared in Experiment 1. Concretely, the results suggest that people might be at least 76% resource-rational. Improving the model further could lead to further upward adjustments of this estimate.

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 performance was affected by the experimental intervention. To address this question, we compared participants' *net* relative performance. In contrast to the *gross*

relative performance measure used in the preceding paragraph and shown in Figure 6 from Experiment 1, this measure includes the cost of gathering information in addition to the payoff of the chosen gamble. 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 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 ($t(100) = 2.60, p = 0.011, d = 0.52$). The difference was not significant in any other environment (LD-LC: $t(102) = 1.51, p = 0.13, d = 0.30$; LD-HC: $t(100) = -1.77, p = 0.079, d = -0.35$; HD-HC: $t(94) = 0.73, p = 0.47, d = 0.15$). Across all conditions, participants in the experimental group were not significantly more resource rational than participants in the control group ($t(402) = 1.12, p = 0.26, d = 0.11$).

Participants in the experimental group should be expected to choose the gamble with the highest subjective expected value on 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, $t(402) = -2.34, p = 0.02, d = -.23$). This counter-intuitive 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 seconds. 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 E for details).

When excluding low-effort participant from both groups, participants in the experimental group were significantly more resource-rational than participants in the

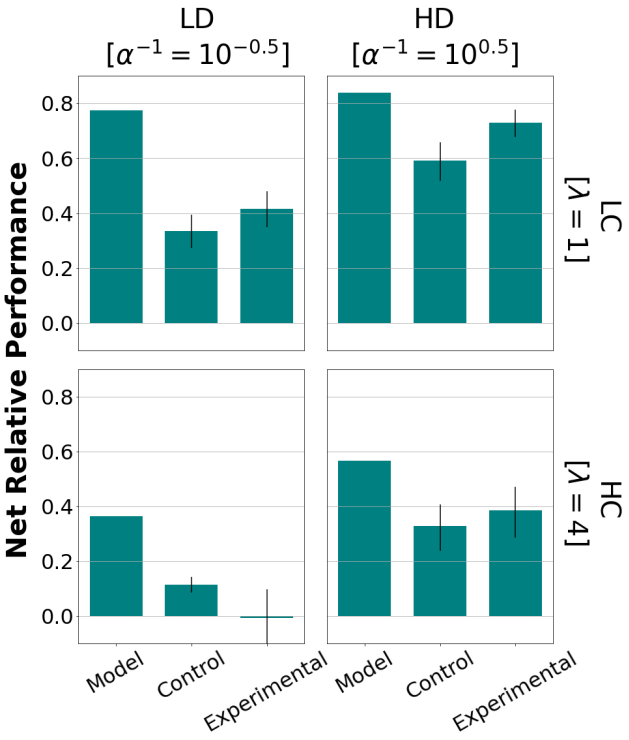

Figure 13

Performance across conditions for each group in Experiment 2. Net relative performance, which accounts for the cost of gathering information, shows that participants in the experimental condition tended to improve performance, but not in all conditions (see Figure E6 for a comparison when excluding low-effort participants). Error-bars show 95% CI across participants.

control group in every condition (LD-LC: $t(55) = 2.36, p = 0.022, d = 0.63$; LD-HC: $t(80) = 4.02, p < 0.001, d = 0.90$; HD-LC: $t(78) = 2.26, p = 0.027, d = 0.51$; HD-HC: $t(73) = 2.30, p = 0.024, d = 0.53$).

Sources of under-performance

As in Experiment 1, we measured participants' net performance and four sources of under-performance as a percentage of the model's net performance (to account for differences across conditions). 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.

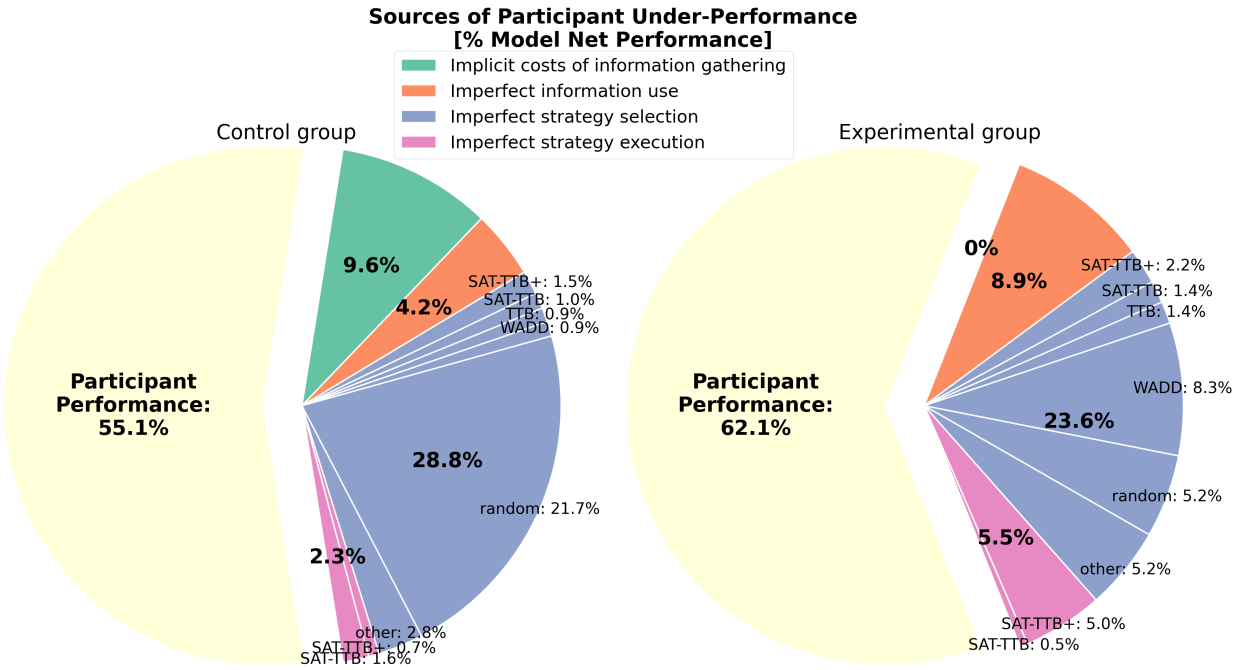

Figure 14

Source of under-performance for each group in Experiment 2. 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 under-performance. 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 E7 shows the same results after excluding low-effort participants from both groups.

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 under-performance 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

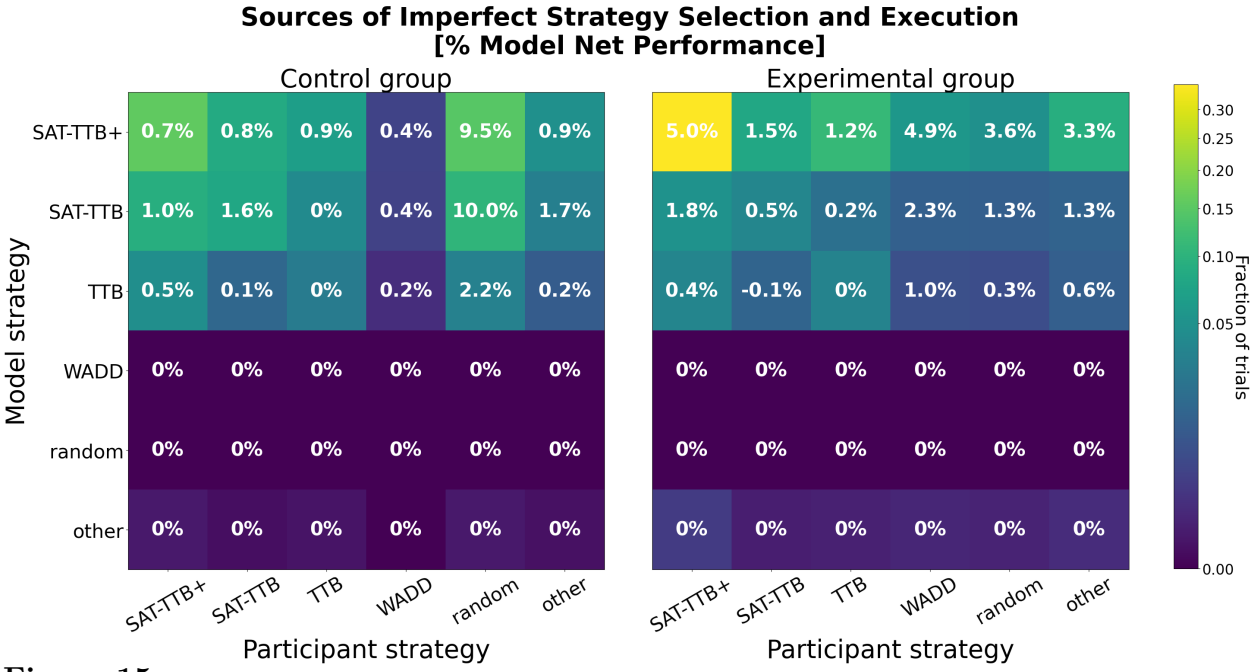

Figure 15

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. Experimental participants' excessive use of WADD occurred mostly when they should have used SAT-TTB+, according to the model, while control participants' excessive use of random gambling occurred mostly when they should have used SAT-TTB.

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 E7, when excluding low-performing participants, imperfect information use accounted for 2.1% of under-performance 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 under-performance in the 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 21.7%) of

under-performance. In the experimental group, this proportion was reduced to 5.2%. On the other hand, whereas the use of WADD accounted for only 0.9% of under-performance 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 under-performance in both groups, it was slightly more in the experimental group, due to the increased usage of the difficult-to-execute SAT-TTB+ 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 WADD in the experimental group, most of it—4.9%—occurred when the best strategy to select was SAT-TTB+.

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 increase in the use of WADD and SAT-TTB+. 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 heuristics 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 heuristics that had previously been overlooked. Furthermore, we collected a large dataset 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 heuristics that our method identified, and they adaptively select which heuristic 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 heuristics and predicting when they should be used as one of finding the optimal policy of a meta-level Markov Decision Process (Hay et al., 2012; Russell & Wefald, 1991b). The meta-level MDP framework allows us to identify those heuristics 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. While information gathering in Mouselab has previously been studied from a resource-rational perspective (Gabaix et al., 2006), the meta-MDP framework provides a new set of computational tools for understanding heuristics through this lens. The result is that we can formally identify heuristics that achieve an optimal trade-off between computational costs and decision quality. By automatically deriving heuristics 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-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-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 heuristics? 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

(Figures 6). Crucially, the under-performance 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 et al., 2022; Consul et al., 2022; Mehta et al., 2022; Skirzyński et al., 2021).

Why did people under-perform 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 under-performance 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 under-performance, 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 finer-grained set of operations is thus an important direction for future work. By expanding the set of computational actions available, we can potentially identify more nuanced strategies and achieve an even closer correspondence to human behavior.

Overall, our findings show that participants use resource-rational heuristics 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.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

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.

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.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of management review*, *3*(3), 439–449.

Becker, F., Skirzynski, J., van Opheusden, B., & Lieder, F. (2022). Boosting human decision-making with ai-generated decision aids. *Computational Brain & Behavior*, *5*(3).

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., et al. (2019). Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *5*, 175–192.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk, 1738 (english translation). *Econometrica*, *22*(1), 23–36. doi: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.

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.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21.

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*, 103896.

Binz, M., Gershman, S. J., Schulz, E., & Endres, D. (2022). Heuristics from bounded meta-learned inference. *Psychological Review*.

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.

Bossaerts, P., & Murawski, C. (2017). Computational complexity and human decision-making. *Trends in Cognitive Sciences*, *21*(12), 917–929.

Bossaerts, P., Yadav, N., & Murawski, C. (2019). Uncertainty and computational complexity. *Philosophical Transactions of the Royal Society B*, *374*(1766), 20180138.

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.

Brown, S., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology*, *53*(6), 453–462.

Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P. M., & Griffiths, T. L. (2018). A resource-rational analysis of human planning. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*.

Callaway, F., Gul, S., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2018). Learning to select computations. *Uncertainty in Artificial Intelligence*.

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*, *119*(12), e2117432119.

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLOS Computational Biology*, *17*(3), e1008863.

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.

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience*, *27*(2), 371–378.

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.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? *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.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological bulletin*, *81*(2), 95–106.

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective and Behavioral Neuroscience*, *8*(4), 429–453.

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.

Dhar, R., Nowlis, S. M., & Sherman, S. J. (1999). Comparison effects on preference construction. *Journal of consumer research*, *26*(3), 293–306.

Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, *35*(7), 1801–1813.

Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, *51*(4), 380–473.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational behavior and human performance*, *13*(2), 171–192.

Elkan, C. (2003). Using the triangle inequality to accelerate k-means. *Proceedings of the 20th international conference on Machine Learning (ICML-03)*, 147–153.

Fishburn, P. C. (1989). Foundations of decision analysis: Along the way. *Management science*, *35*(4), 387–405.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, *7*(2), 179–188.

Frank, M. C. (2013). Throwing out the Bayesian baby with the optimal bathwater: Response to. *Cognition*, *128*(3), 417–423.

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.

Gardner, J. L. (2019). Optimality and heuristics in perceptual neuroscience. *Nature neuroscience*, *22*(4), 514–523.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267–314.

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.

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.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, *62*, 451–482.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. *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, USA.

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*, *108*(Supplement 3), 15647–15654.

Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of experimental psychology: Learning, memory, and cognition*, *34*(5), 1055.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90.

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.

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.

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.

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

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*. AUAI Press.

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).

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.

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.

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*(3), 733–758.

Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine learning*, *11*(1), 63–90.

Howard, R. A. (1968). The foundations of decision analysis. *IEEE transactions on systems science and cybernetics*, *4*(3), 211–219.

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.

Huygens, C. (1657). *De ratiociniis in ludo aleae*. 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 (english translation)*. 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.

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management science*, *31*(4), 395–414.

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.

Katsikopoulos, K. V., & Martignon, L. (2006). Naive heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, *50*(5), 488–494.

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.

Klein, G. (2008). Naturalistic decision making. *Human factors*, *50*(3), 456–460.

Kwisthout, J., Wareham, T., & van Rooij, I. (2011). Bayesian intractability is not an ailment that approximation can cure. *Cogn. Sci.*, *35*(5), 779–784.

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.

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.

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.

Lichtenberg, J. M., & Şimşek, Ö. (2017). Simple regression models. *Imperfect decision makers: Admitting real-world rationality*, 13–25.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762–794.

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*.

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.

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

Manzini, P., & Mariotti, M. (2012). Choice by lexicographic semiorders. *Theoretical Economics*, *7*(1), 1–23.

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

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

Martignon, L., Hoffrage, U., Group, A. R., et al. (1999). Why does one-reason decision making work. *Simple heuristics that make us smart*, 119–140.

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.

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*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

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

Newell, A., Simon, H. A. et al. (1972). *Human problem solving* (Vol. 104). Prentice-hall Englewood Cliffs, NJ.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154.

Nowozin, S. (2014). Optimal decisions from probabilistic models: The intersection-over-union case. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 548–555.

Papadimitriou, C. H., & Tsitsiklis, J. (1986). Intractable problems in control theory. *SIAM journal on control and optimization*, *24*(4), 639–654.

Parpart, P., Jones, M., & Love, B. C. (2018). Heuristics as bayesian inference under extreme priors. *Cognitive psychology*, *102*, 127–144.

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.

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.

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*, 2825–2830.

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

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. 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.

Reisen, N., Hoffrage, U., & Mast, F. W. (2008). Identifying decision strategies in a consumer choice situation. *Judgment and decision making*, *3*(8), 641–658.

Rescorla, R. A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, 64–99.

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.

RStudio Team. (2019). *Rstudio: Integrated development environment for r*. RStudio, Inc. Boston, MA. http://www.rstudio.com/

Russell, S. (2016). Rationality and Intelligence : A Brief Update. In Müller V. C. (Ed.), *Fundamental Issues of Artificial Intelligence* (pp. 1–21).

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.

Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *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.

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical association*, *46*(253), 55–67.

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.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.

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.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological bulletin*, *134*(2), 207–222.

Shteingart, H., & Loewenstein, Y. (2014). Reinforcement learning and human behavior. *Current Opinion in Neurobiology*, *25*, 93–98.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of go without human knowledge. *nature*, *550*(7676), 354–359.

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. *Decision and organization*, *1*(1), 161–176.

Simon, H. A. (1990). Invariants of human behavior. *Annual review of psychology*, *41*(1), 1–20.

Şimşek, Ö. (2013). Linear decision rule as aspiration for simple decision heuristics. *Advances in neural information processing systems*, *26*, 2904–2912.

Şimşek, Ö., & Buckmann, M. (2015). Learning from small samples: An analysis of simple decision heuristics. *Advances in neural information processing systems*, *28*, 3159–3167.

Skirzyński, J., Becker, F., & Lieder, F. (2021). Automatic discovery of interpretable planning strategies. *Machine Learning*, 2641–2683.

Stigler, G. J. (1961). The economics of information. *Journal of political economy*, *69*(3), 213–225.

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.

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.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*(3), 219–225.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*(1), 31–48.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

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., ... 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. *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.

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.

Zednik, C., & Jäkel, F. (2016). Bayesian reverse-engineering considered as a research strategy for cognitive science. *Synthese*, *193*(12), 3951–3985.

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, Gul, 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

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

subject to the constraints that $w_1, \cdots, w_3 \in [0, 1]$, $w_1 + w_2 + w_3 = 1$, 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

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

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

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

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,

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

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

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

We applied BMPS separately to each of the fifty meta-level MDPs modelling the fifty 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 highest expected performance. See Callaway, Gul, et al. (2018) for details of the BMPS optimization procedure.

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.

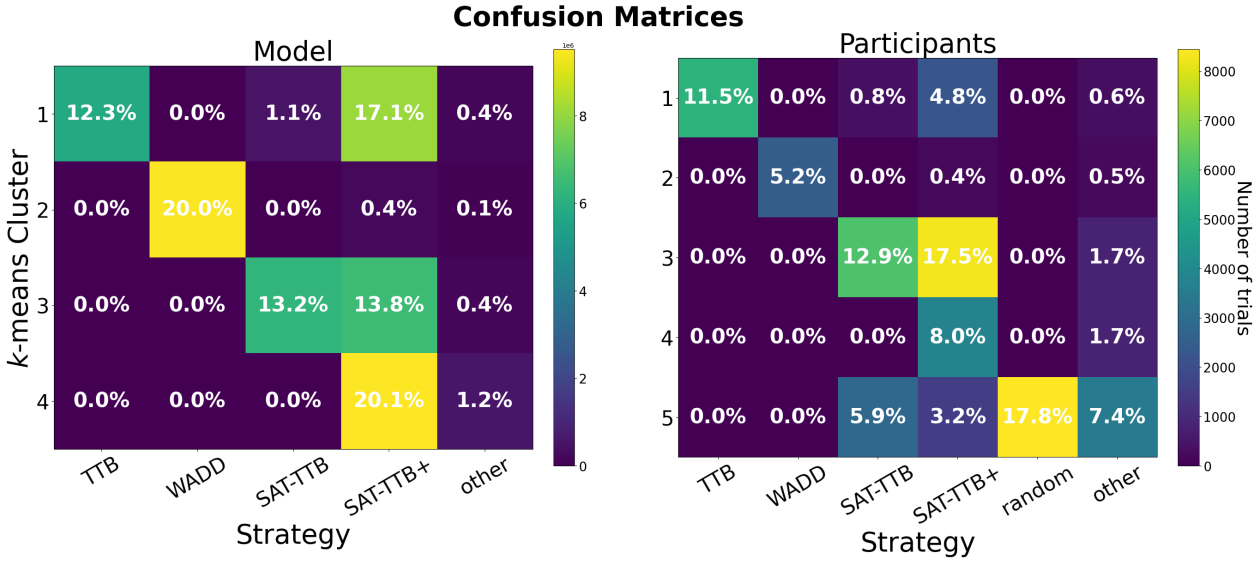

Figure B1

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. 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.

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.

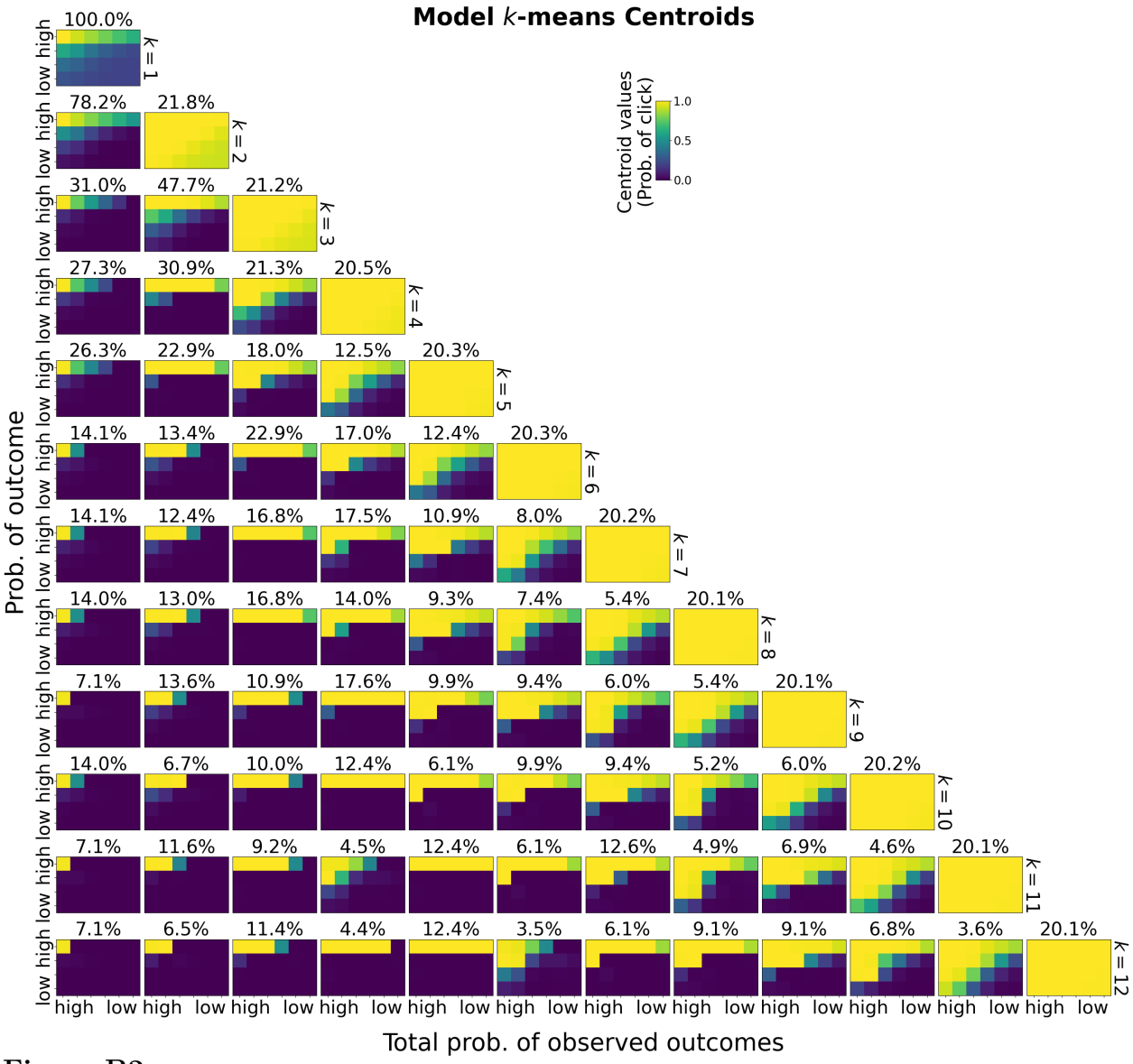

Figure B2

k-means clustering results for model data in Experiment 1. 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.*

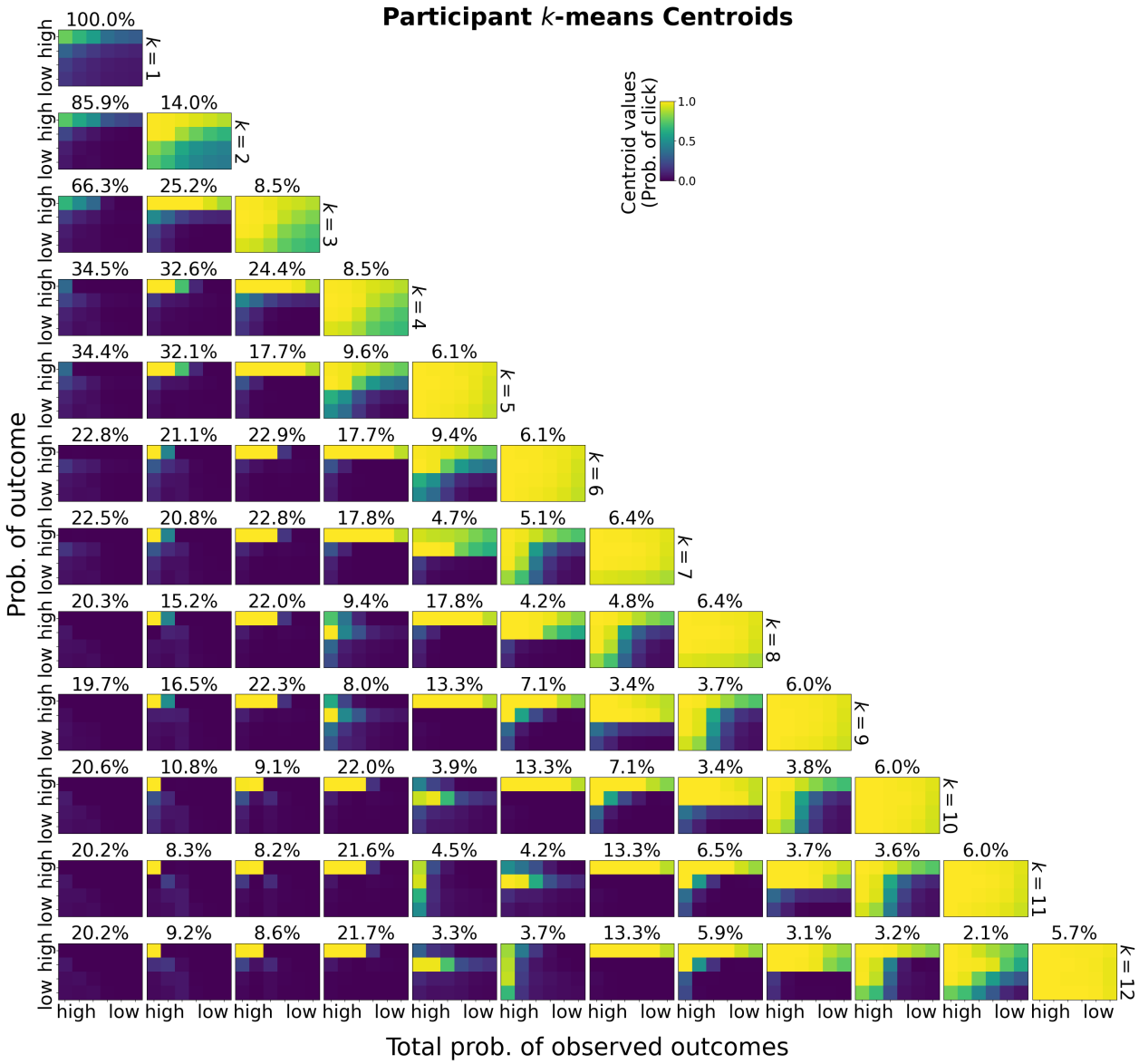

Figure B3

Same as previous figure, but with participant data from Experiment 1.

Appendix C

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 fifty environments (with 2 levels of stakes $\times$ 5 levels of dispersion $\times$ 5 levels of cost). They illustrate overall qualitative correspondence between the model and participants in adaptive application of strategies according to the statistics of the environment.

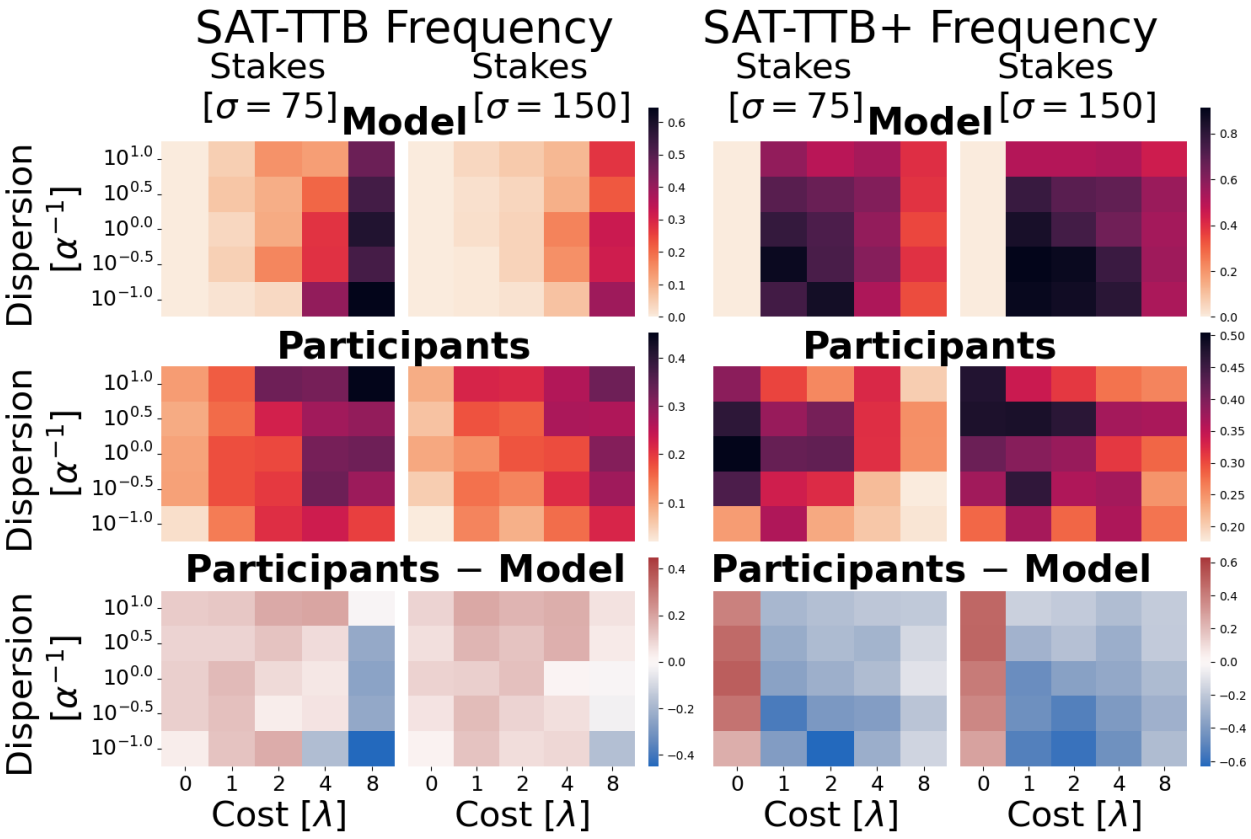

Figure C1

Frequency of SAT-TTB (left panels) and SAT-TTB+ (right panels) across all fifty experimental conditions, for the model (top panels), participants (middle panels), and a comparison between the model and participants (bottom panels) from Experiment 1. 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. SAT-TTB+ and SAT-TTB are two heuristics discovered using our resource-rational method.

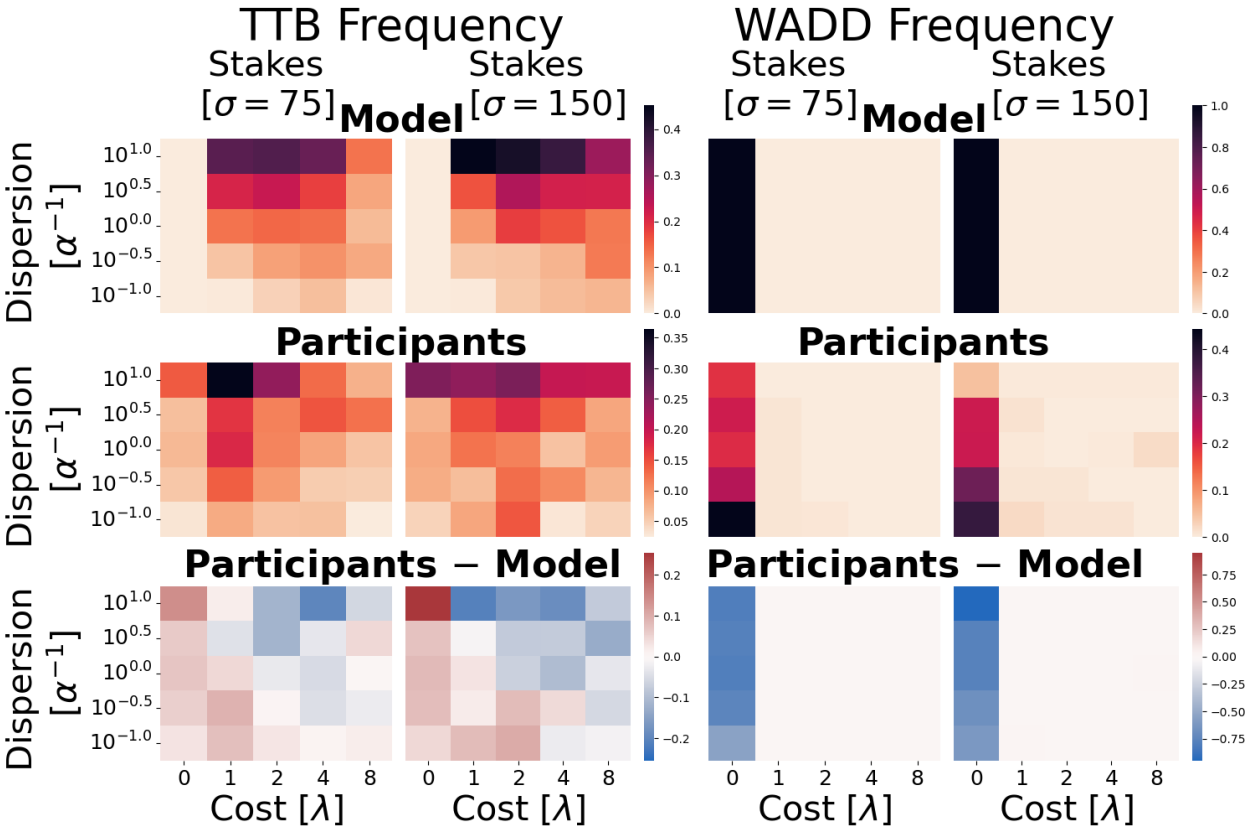

Figure C2

TTB (left panels) and WADD (right panels) strategy use frequencies across all fifty 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. TTB and WADD 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.

Table C1

Statistical results accompanying Figure 4 from Experiment 1.

Strategy	Independent variable	significant post-hoc comparisons	effect sizes (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&8	$-0.21, 0.047, 0.13, 0.063$
SAT-TTB	cost	all pairs	$-0.27, -0.078, -0.16, -0.089$

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).

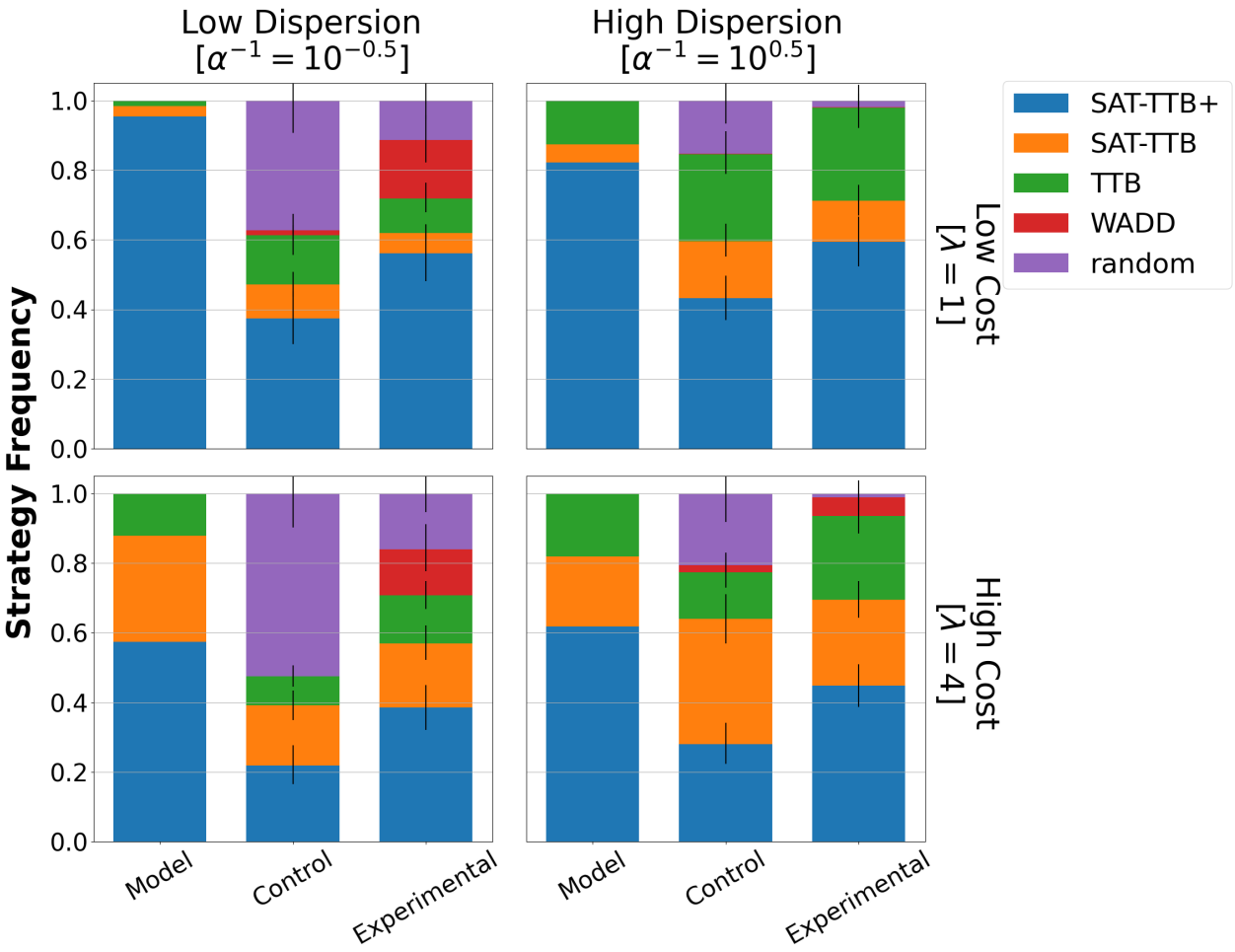

Figure C3

Reducing implicit costs increases the use of costly heuristics and reduces random gambling for participants in the experimental group in from Experiment 2. Compare with Figure 11, which omits random gambling.

Experiment 2

To facilitate comparison with Experiment 1 (Figure 4) in the main text, Figure 11 is conditioned on the same four strategies (that is, omitting random gambling and unidentified patterns of clicking). Figure C3 includes random gambling to illustrate how much this decreased in the experimental group compared to the control group. Figure C4 includes all trials.

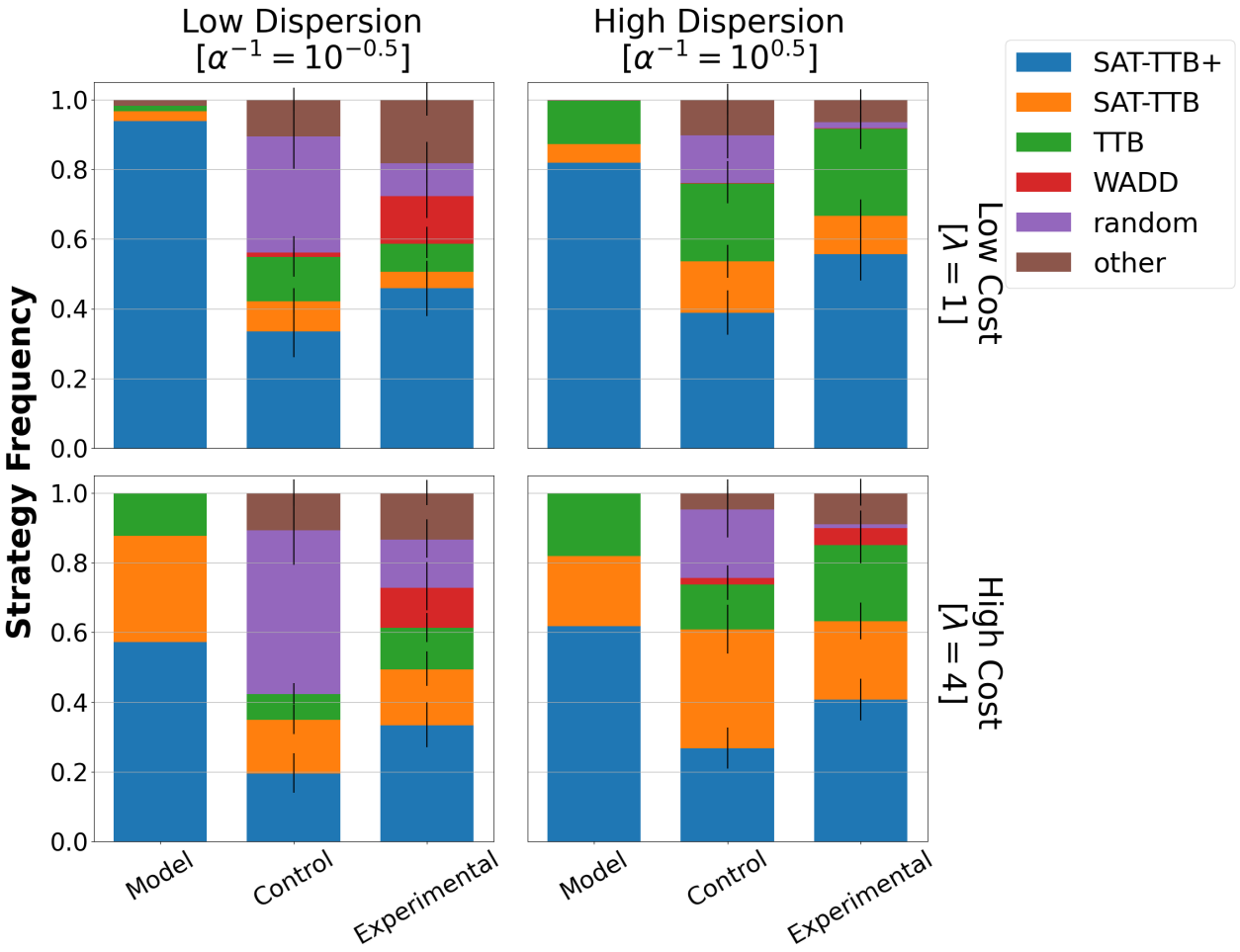

Figure C4

Same as previous figure but also including unidentified patterns of clicking from Experiment 2.

Appendix D

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 heuristics 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 D1 displays these two variables in all fifty environmental conditions. Figures D2 and D3 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 subsection *Rational strategy selection explains variability in choice behavior* in the Results section of the main text for details on how these measurements were defined.

Table D1 summarizes statistical analyses accompanying those presented in the main text, corresponding to Figures 5, D2, and 6. 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-HSD correction for multiple comparisons. The effect sizes for these comparisons were calculated using Cohen's d.

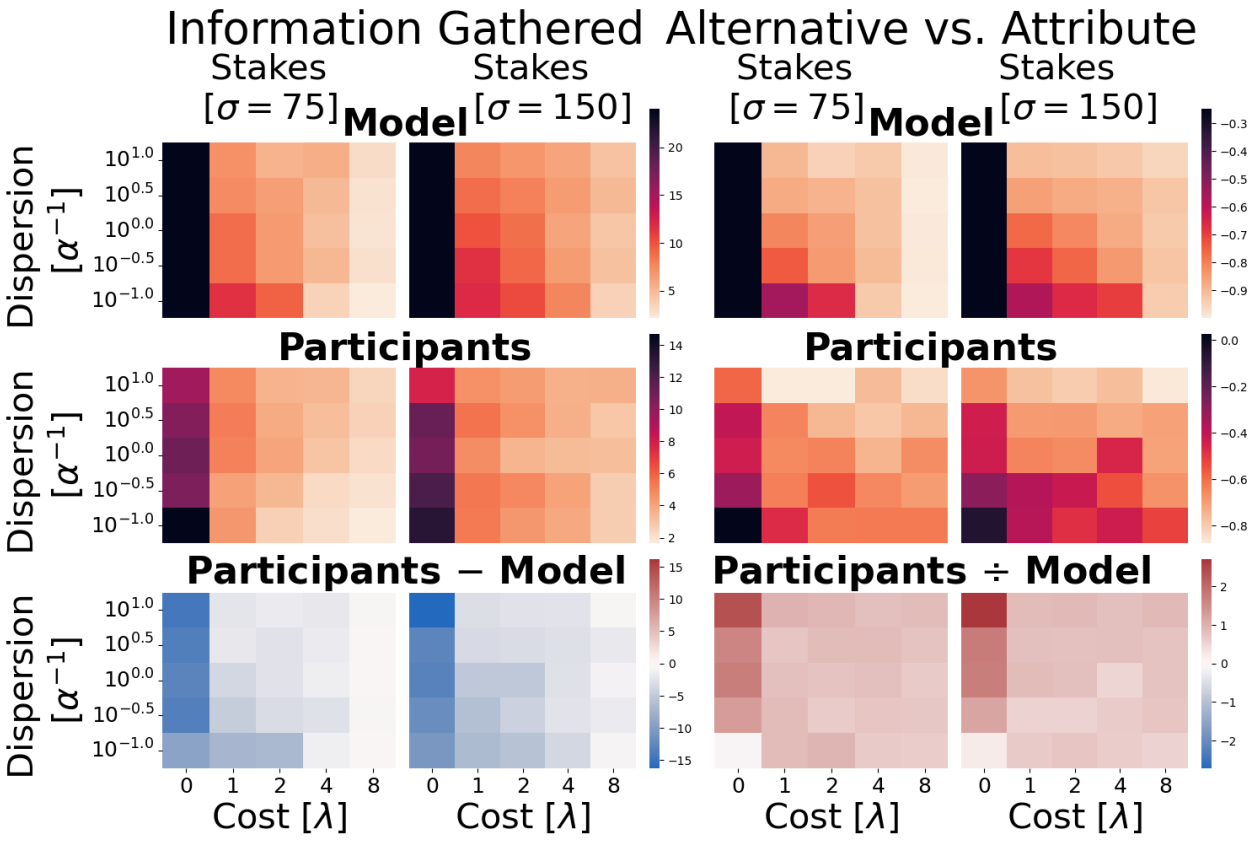

Figure D1

Information-gathering (measured with clicks; left panels) and attribute- versus alternative-based processing (right panels) shown across all fifty conditions of Experiment 1, for the model (top row), human participants (middle row), and a comparison between the model and participants (bottom row). The fifty conditions vary three parameters for a 2x5x5 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 fifty conditions is shown. See the subsection Rational strategy selection explains variability in choice behavior in the Results section of the main text for details on how alternative- versus attribute-based processing was measured.

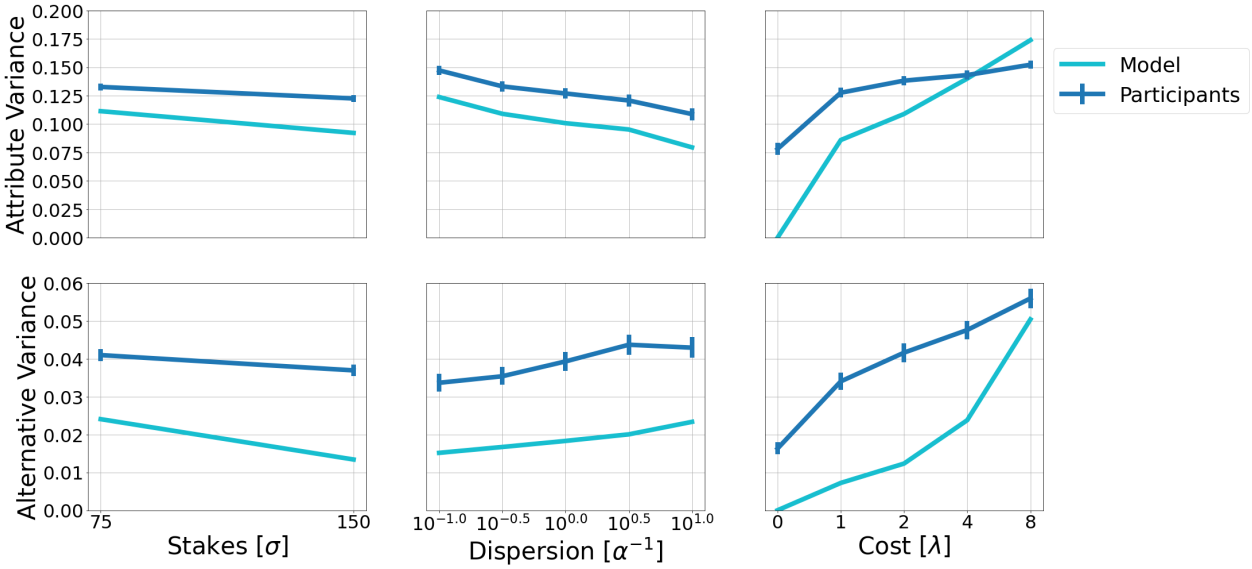

Figure D2

Behavioral correspondence between participants and the resource-rational model from Experiment 1. 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% CI across participants.

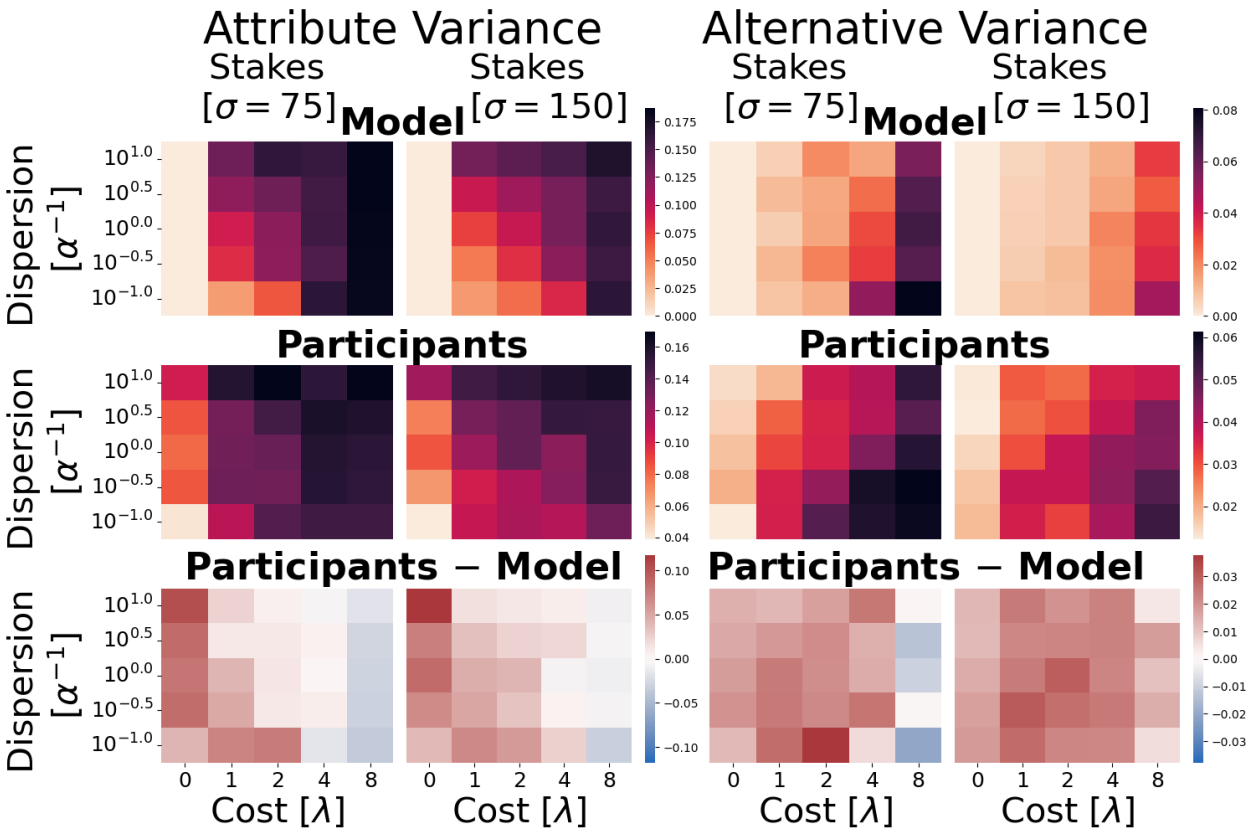

Figure D3

Alternative and attribute variance for all fifty 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). The results here accompany the behavioral results shown in Figure 5. Within each parameter value in Figure fig:clicks_processing, results are averaged across all values of other parameters, whereas in this figure the full results for each of the fifty conditions is shown.

Table D1

Statistical results accompanying Figures 5 and D2 from Experiment 1.

Behavioral feature	Independent variable	main effect	significant post-hoc comparisons	effect sizes (Cohen's d)
Information gathering	stakes	$t(2366) = -2.61$, $p = 0.009$	n/a	-0.11
Information gathering	dispersion	$F(4,2363) = 1.22$, $p = 0.3$	n/a	$0.064, 0.0012, -0.036, 0.11$
Information gathering	cost	$F(4,2363) = 293.83$, $p < 0.001$	all pairs except 2&4, 4&8	$1.0, 0.32, 0.25, 0.29$
Alternative vs. Attribute	stakes	$t(2131) = -2.28$, $p = 0.022$	n/a	-0.099
Alternative vs. Attribute	dispersion	$F(4,2128) = 27.97$, $p < 0.001$	all pairs except $10^{-1.0}$&$10^{-0.5}$, 10^{0}&$10^{0.5}$	$0.16, 0.2, 0.092, 0.23$
Alternative vs. Attribute	cost	$F(4,2128) = 31.44$, $p < 0.001$	0&1, 0&2, 0&4, 0&8	$0.52, 0.048, -0.012, 0.12$
Attribute variance	stakes	$t(2195) = 3.89$, $p < 0.001$	n/a	0.17
Attribute variance	dispersion	$F(4,2192) = 24.74$, $p < 0.001$	all pairs except $10^{-0.5}$&10^{0}, 10^{0}&$10^{0.5}$	$-0.18, -0.1, -0.11, -0.26$
Attribute variance	cost	$F(4,2192) = 121.75$, $p < 0.001$	all pairs except 2&4, 4&8	$-0.78, -0.2, -0.095, -0.19$
Alternative variance	stakes	$t(2195) = 2.93$, $p = 0.0035$	n/a	0.12
Alternative variance	dispersion	$F(4,2192) = 8.43$, $p < 0.001$	$10^{-1.0}$&$10^{0.5}$, $10^{-1.0}$&$10^{1.0}$, $10^{-0.5}$&$10^{0.5}$, $10^{-0.5}$&$10^{1.0}$	$-0.023, 0.14, 0.13, 0.057$
Alternative variance	cost	$F(4,2192) = 115.01$, $p < 0.001$	all pairs	$-0.7, -0.24, -0.19, -0.26$

Summary of statistical results corresponding to the analyses shown in Figures 5 and D2 from Experiment 1. A two-sample t-test was used to test the main effect of stakes on the dependent variables. ANOVAs 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-HSD 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).

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 D4, the pattern of results reflects the overall increase in information gathering in the experimental group: decreased attribute-variance and alternative-variance (Figure D4B and C), 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 D4A). The statistical results of comparing these measurements across the experimental group and the control group are summarized in Table D2, and similar results comparing these measurements between the model and each group are presented in Table D3, 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.

Table D2

Statistical results accompanying Figure D4 from Experiment 2.

Behavioral feature	Condition (dispersion, cost)	t-statistic	p-value	effect size (Cohen's d)
Processing pattern	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 1$	$t(90) = 1.39$	$p = 0.17$	$d = 0.29$
Processing pattern	$\alpha^{-1} = 10^{0.5}$ $\lambda = 1$	$t(95) = -1.37$	$p = 0.17$	$d = -0.28$
Processing pattern	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 4$	$t(85) = 2.94$	$p = 0.0042$	$d = 0.64$
Processing pattern	$\alpha^{-1} = 10^{0.5}$ $\lambda = 4$	$t(86) = 3.43$	$p < 0.001$	$d = 0.73$
Attribute variance	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 1$	$t(92) = -3.23$	$p = 0.0017$	$d = -0.67$
Attribute variance	$\alpha^{-1} = 10^{0.5}$ $\lambda = 1$	$t(96) = -0.94$	$p = 0.35$	$d = -0.19$
Attribute variance	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 4$	$t(88) = -3.16$	$p = 0.0021$	$d = -0.67$
Attribute variance	$\alpha^{-1} = 10^{0.5}$ $\lambda = 4$	$t(90) = -3.73$	$p < 0.001$	$d = -0.78$
Alternative variance	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 1$	$t(92) = -2.24$	$p = 0.027$	$d = -0.46$
Alternative variance	$\alpha^{-1} = 10^{0.5}$ $\lambda = 1$	$t(96) = -2.02$	$p = 0.046$	$d = -0.41$
Alternative variance	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 4$	$t(88) = -2.21$	$p = 0.03$	$d = -0.47$
Alternative variance	$\alpha^{-1} = 10^{0.5}$ $\lambda = 4$	$t(90) = -3.54$	$p < 0.001$	$d = -0.74$

Summary of comparisons between the experimental group and the control group for the behavioral measures shown in Figure D4 from Experiment 2.

Table D3

Statistical results accompanying Figure D4 from Experiment 2.

Behavioral feature	Condition (dispersion, cost)	t-statistic	p-value	effect size (Cohen's d)
Processing pattern	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 1$	$t(90) = 0.94$	$p = 0.35$	$d = 0.20$
Processing pattern	$\alpha^{-1} = 10^{0.5}$ $\lambda = 1$	$t(95) = -1.50$	$p = 0.14$	$d = -0.31$
Processing pattern	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 4$	$t(85) = 2.74$	$p = 0.0076$	$d = 0.59$
Processing pattern	$\alpha^{-1} = 10^{0.5}$ $\lambda = 4$	$t(86) = 2.91$	$p = 0.0046$	$d = 0.62$
Attribute variance	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 1$	$t(92) = -1.25$	$p = 0.21$	$d = -0.26$
Attribute variance	$\alpha^{-1} = 10^{0.5}$ $\lambda = 1$	$t(96) = -0.56$	$p = 0.58$	$d = -0.11$
Attribute variance	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 4$	$t(88) = 2.83$	$p = 0.0058$	$d = 0.60$
Attribute variance	$\alpha^{-1} = 10^{0.5}$ $\lambda = 4$	$t(90) = 1.30$	$p = 0.2$	$d = 0.27$
Alternative variance	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 1$	$t(92) = -1.81$	$p = 0.073$	$d = -0.37$
Alternative variance	$\alpha^{-1} = 10^{0.5}$ $\lambda = 1$	$t(96) = -2.08$	$p = 0.04$	$d = -0.42$
Alternative variance	$\alpha^{-1} = 10^{-0.5}$ $\lambda = 4$	$t(88) = 0.03$	$p = 0.97$	$d = 0.01$
Alternative variance	$\alpha^{-1} = 10^{0.5}$ $\lambda = 4$	$t(90) = -3.11$	$p = 0.0025$	$d = -0.65$

Summary of comparisons between each group and the model for the behavioral measures shown in Figure D4 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.

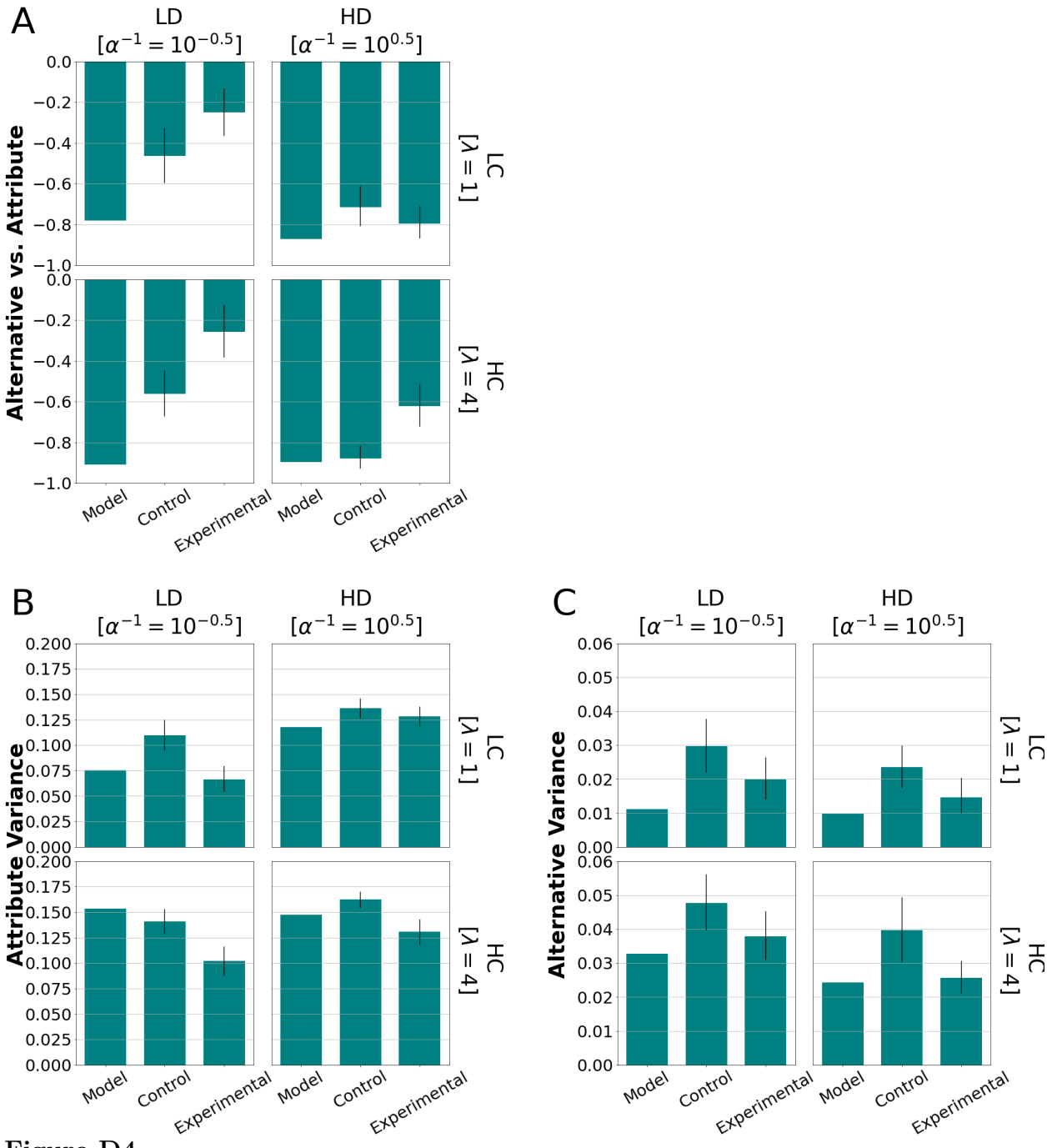

Figure D4

Behavioral features of information processing from Experiment 2. ***(A)*** *Consistent with their over-use of WADD, participants in the experimental condition showed an increase in alternative vs. 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.*

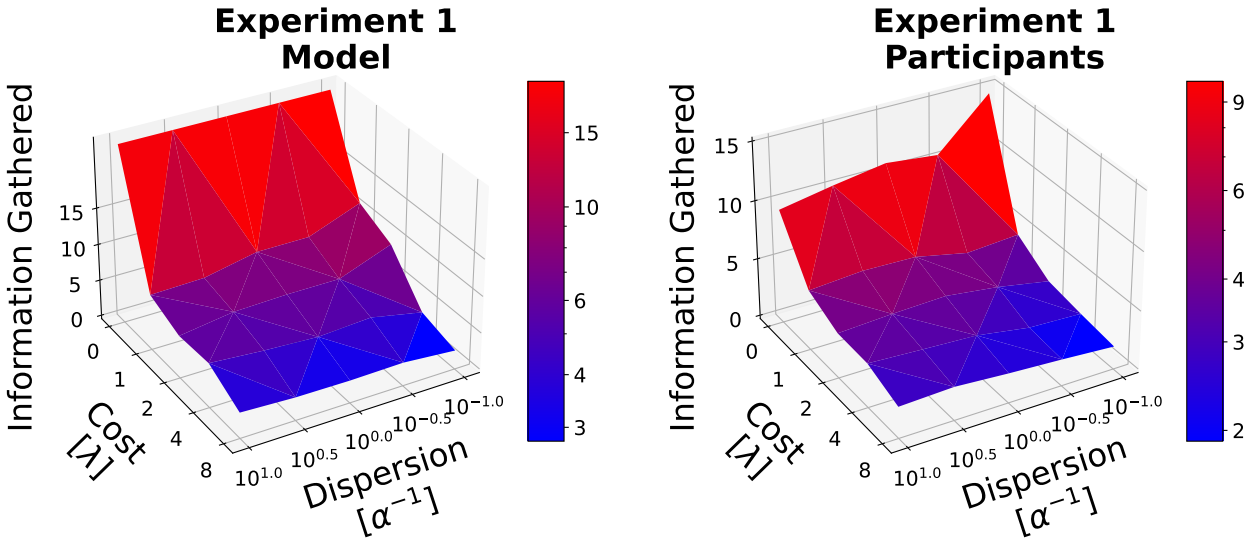

Figure D5

Interaction between cost and dispersion on information gathering. This figure offers a 3D perspective on Figure D6, showing that information gathering decreases with stakes for low cost, but increases with stakes for high cost. This include the low stakes condition only from Experiment 1, for comparison with Experiment 2.

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 D6A. 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 ($t(99) = 3.19, p = 0.0019, d = 0.63$), but unlike the model, for high cost, participants in this group click slightly *less* with high dispersion ($t(95) = 0.40, p = 0.69, d = 0.08$). The opposite pattern holds for the control group: clicking increases with dispersion for both high cost (as predicted by the model; $t(99) = 2.32, p = 0.022, d = 46$) and low cost (unlike the model; $t(103) = 0.63, p = 0.53, d = 0.12$).

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 D6A). 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

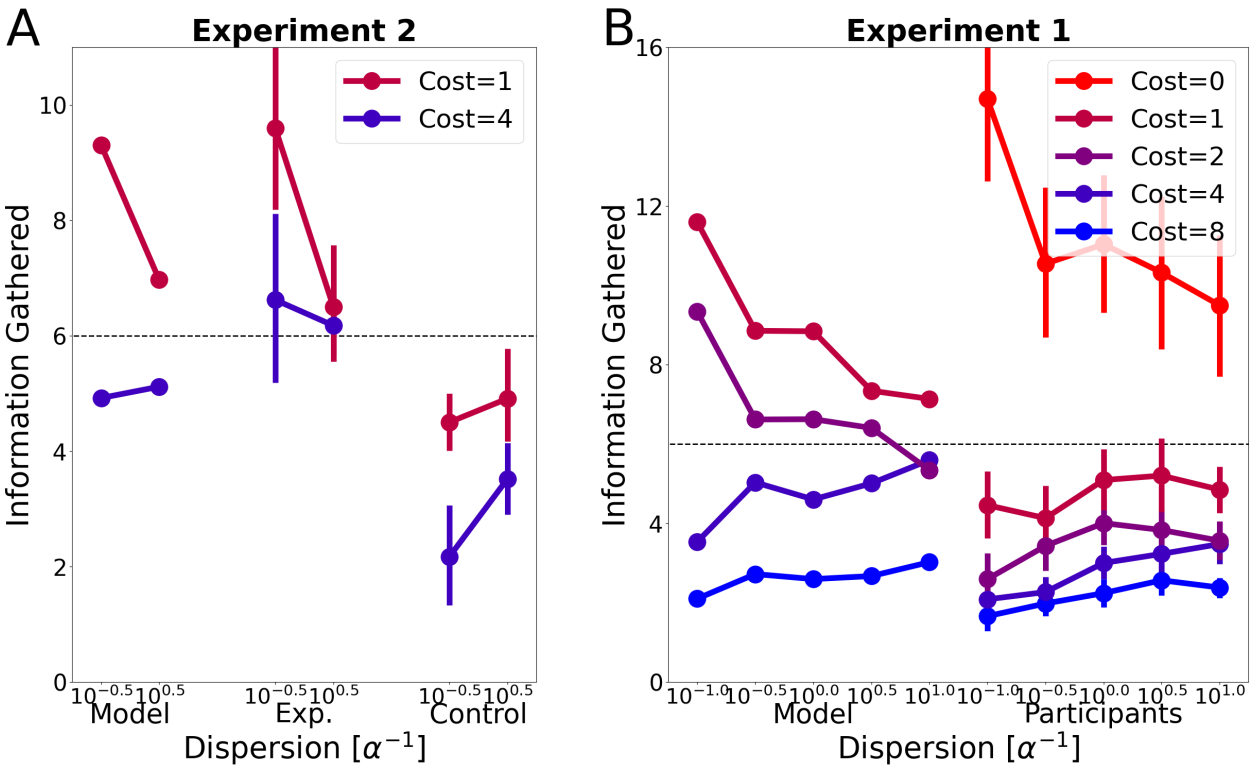

Figure D6

Interaction between cost and dispersion on information gathering. ***(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% CI across participants.*

data from Experiment 1, with five levels of dispersion and cost. As shown in D6B, 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 D5 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 D4). While participants tend to under-perform 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.

Appendix E

Performance and Sources of under-performance

This Appendix provides additional details to accompany the sections on *Performance* and *Sources of under-performance* for each experiment in the main text.

Experiment 1

Performance

Here we provide additional statistical results and figures that show performance when excluding low-effort participants, and performance across all 50 conditions.

Because group under-performance may be driven by low-effort participants who simply do not perform the task, we measured relative performance after excluding participants who gambled randomly on more than half of all trials ($n = 394$ or 16.6% of participants). As illustrated in Figure E1, the average relative performance 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 relative performance on the trials of attentive participants (0.907) was very similar to its relative performance 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 are 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 under-performance 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: $B = 0.029, p = 0.0096$,

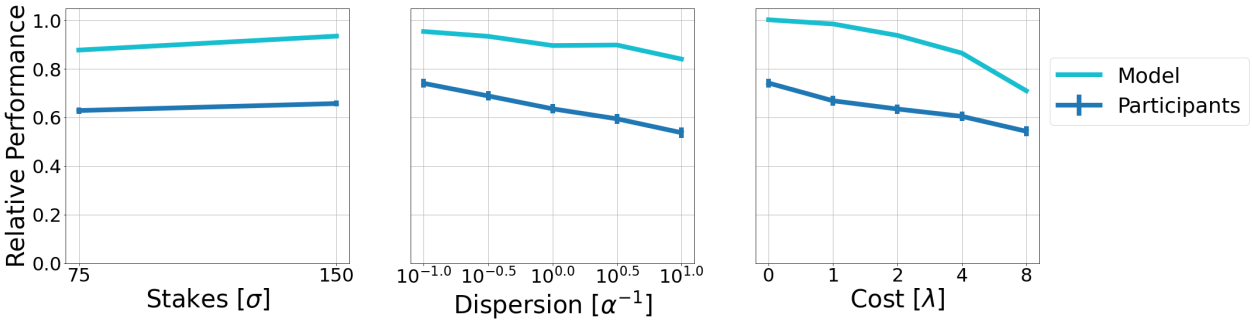

Figure E1

Performance when excluding participants who gamble randomly on more than half of all trials from Experiment 1. 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% CI across participants.

$B = 0.05, p < 0.001$, and $B = -0.046, p < 0.001$, respectively.

Table E1

Statistical results accompanying Figure 6 from Experiment 1.

Behavioral feature	Independent variable	main effect	significant post-hoc comparisons	effect sizes (Cohen's d)
Relative performance	stakes	$t(2366) = \text{-}2.92$, $p = 0.0036$	n/a	-0.12
Relative performance	dispersion	$F(4,2363) = 42.76$, $p < 0.001$	all pairs except $10^{-0.5}$&10^{0}	$-0.22, -0.11, -0.24, -0.19$
Relative performance	cost	$F(4,2363) = 50.48$, $p < 0.001$	all pairs except 1&2, 2&4	$0.36, 0.16, 0.13, 0.2$
Relative performance (with exclusions)	stakes	$t(1972) = \text{-}2.59$, $p = 0.0096$	n/a	-0.12
Relative performance (with exclusions)	dispersion	$F(4,1969) = 43.29$, $p < 0.001$	all pairs except $10^{-0.5}$&10^{0}	$-0.23, -0.17, -0.22, -0.22$
Relative performance (with exclusions)	cost	$F(4,1969) = 38.00$, $p < 0.001$	all pairs except 1&2, 2&4	$0.3, 0.14, 0.13, 0.26$

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. ANOVAs 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-HSD 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).

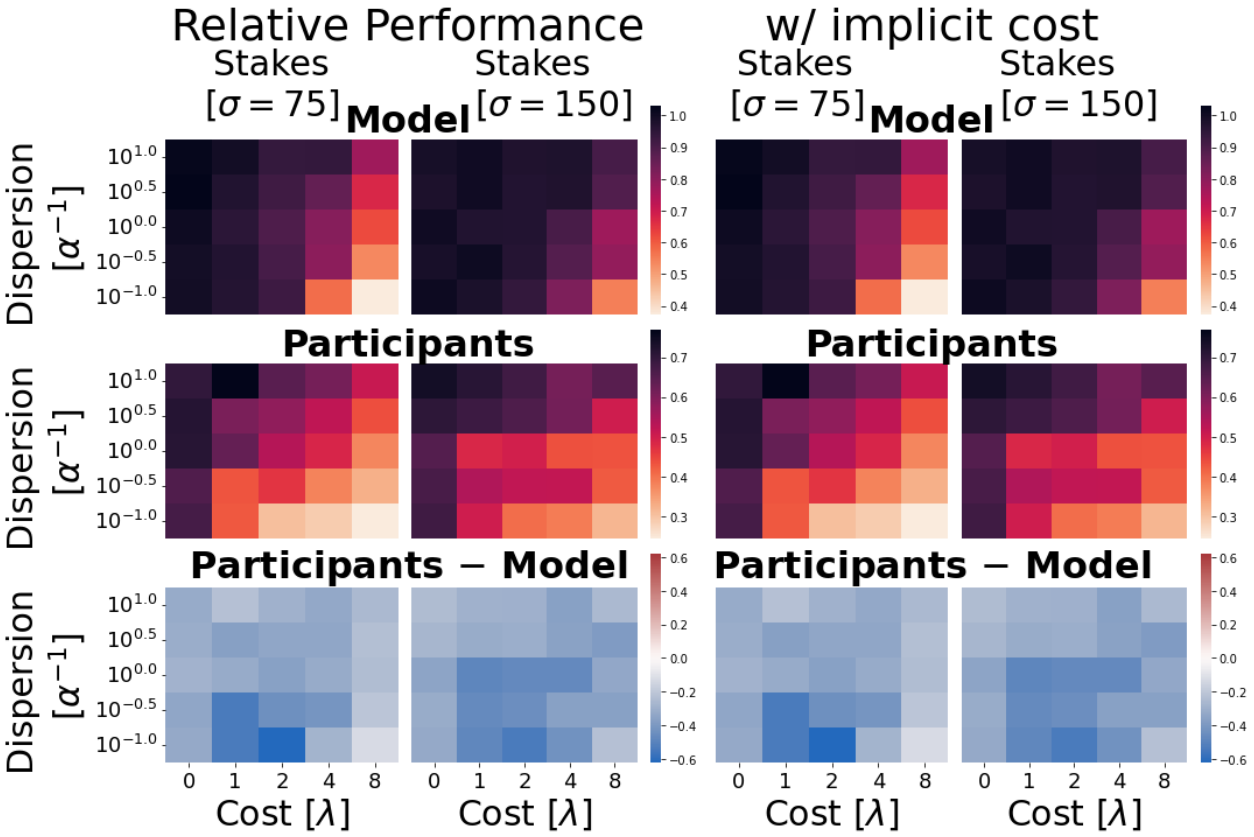

Figure E2

Relative performance (left panels) and relative performance for the model with an implicit cost of clicking (right panels) shown across all fifty conditions of Experiment 1, for the model (top row), human participants (middle row), and the difference between the model and participants (bottom row). The fifty conditions vary three parameters for a 2x5x5 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 S8, results are averaged across all values of other parameters, whereas in this figure the full results for each of the fifty conditions is shown.

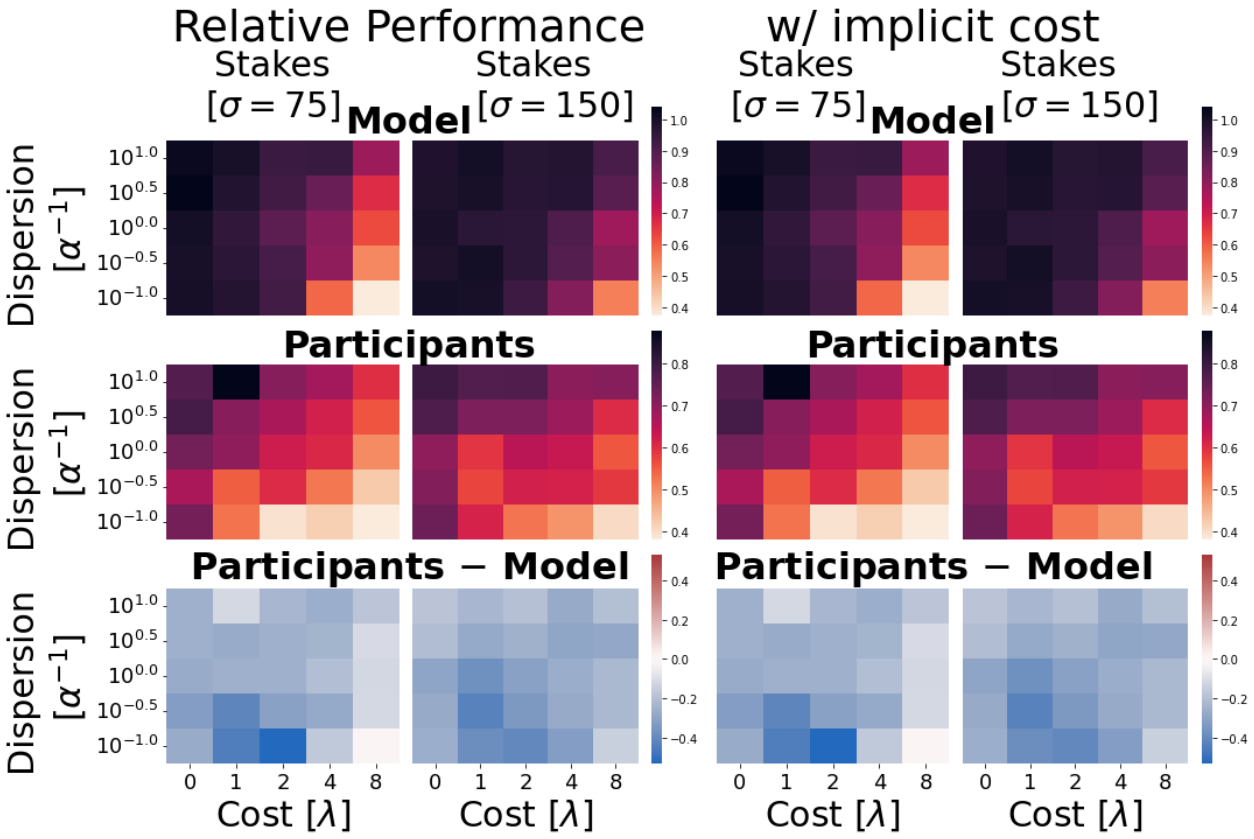

Figure E3

Same as Figure E2, but excluding participants who gambled randomly on more than half of all trials (n = 394 of 2,368 participants total) from Experiment 1.

Sources of under-performance

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

Figure E1 shows that, when excluding low-effort participants, the remaining participants achieved 70.9% of the 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 under-performance: implicit costs of information gathering, imperfect use of the gathered information, imperfect strategy selection, and imperfect strategy execution. As shown in Figure E4, participants achieved 70.6% (95% CI [68.3, 71.2]) of the net performance of the model, with the four sources of under-performance 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 E4). 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 under-performance, but not as high as without participant exclusions (compare to Figure 7). Figures E4 and E5 show the same analyses presented in Figures 7 and 8, respectively, when excluding participants.

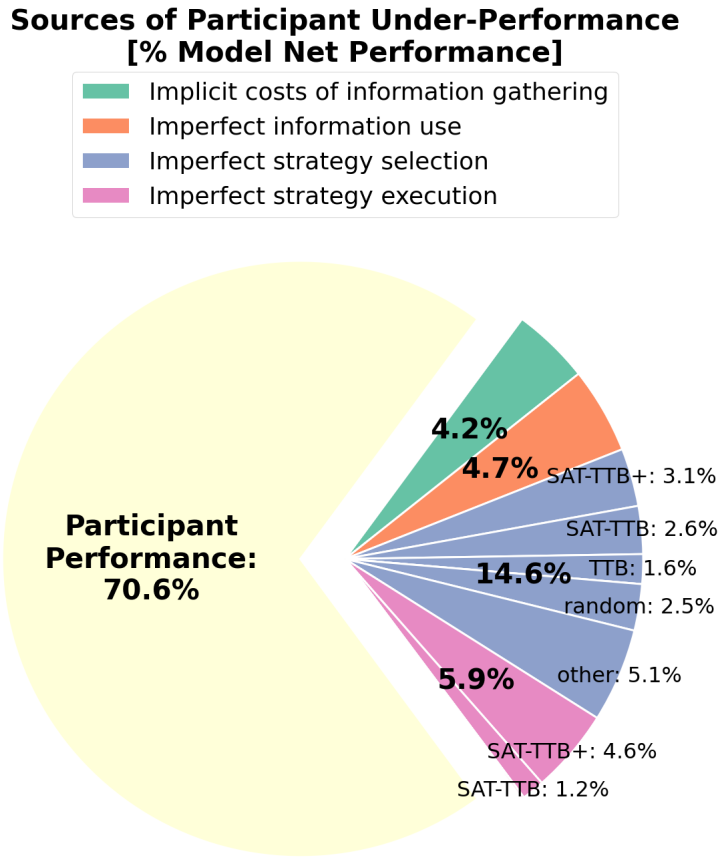

Figure E4

Sources of under-performance when excluding low-effort participants from Experiment 1. 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.*

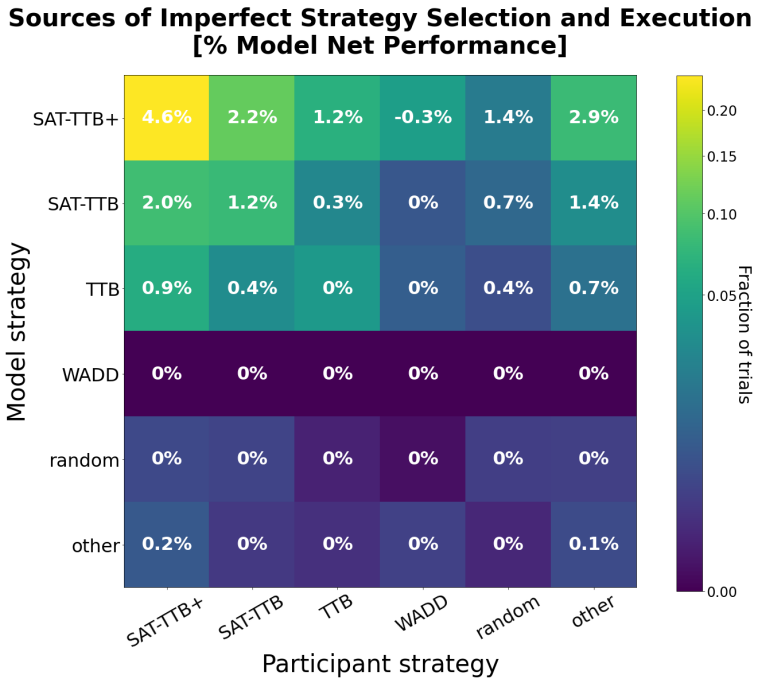

Figure E5

Sources of imperfect strategy selection and execution when excluding low-effort participants from Experiment 1. 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.

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 20s 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 E6 shows that attentive participants in the experimental group out-performed attentive participants in the control group in every condition (LD-LC: $t(55) = 2.36, p = 0.022, d = 0.63$; LD-HC: $t(80) = 4.02, p < 0.001, d = 0.90$; HD-LC: $t(78) = 2.26, p = 0.027, d = 0.51$; HD-HC: $t(73) = 2.30, p = 0.024, d = 0.53$).

Sources of under-performance

We computed the same four sources of under-performance after excluding low-effort participants from both groups. These results are shown in Figures E7 and E8. 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.

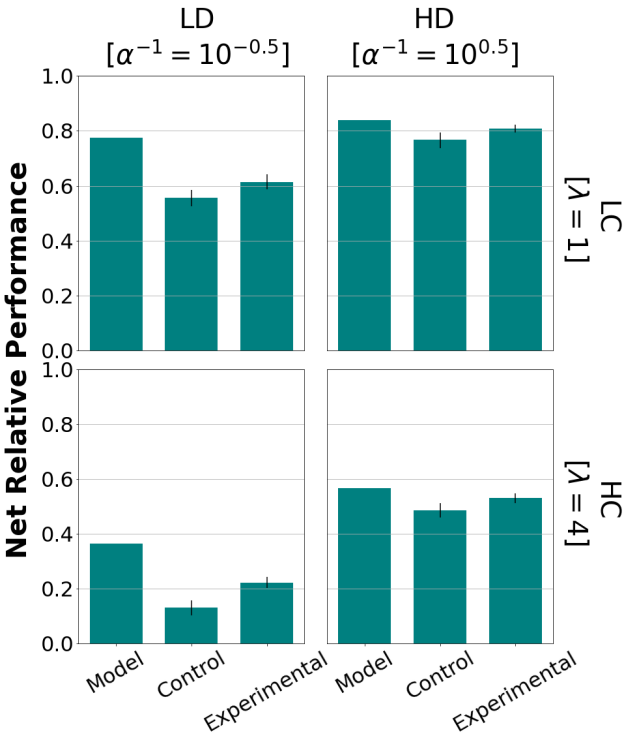

Figure E6

Performance across conditions for each group in Experiment 2 when excluding low-effort participants. 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% CI across participants.

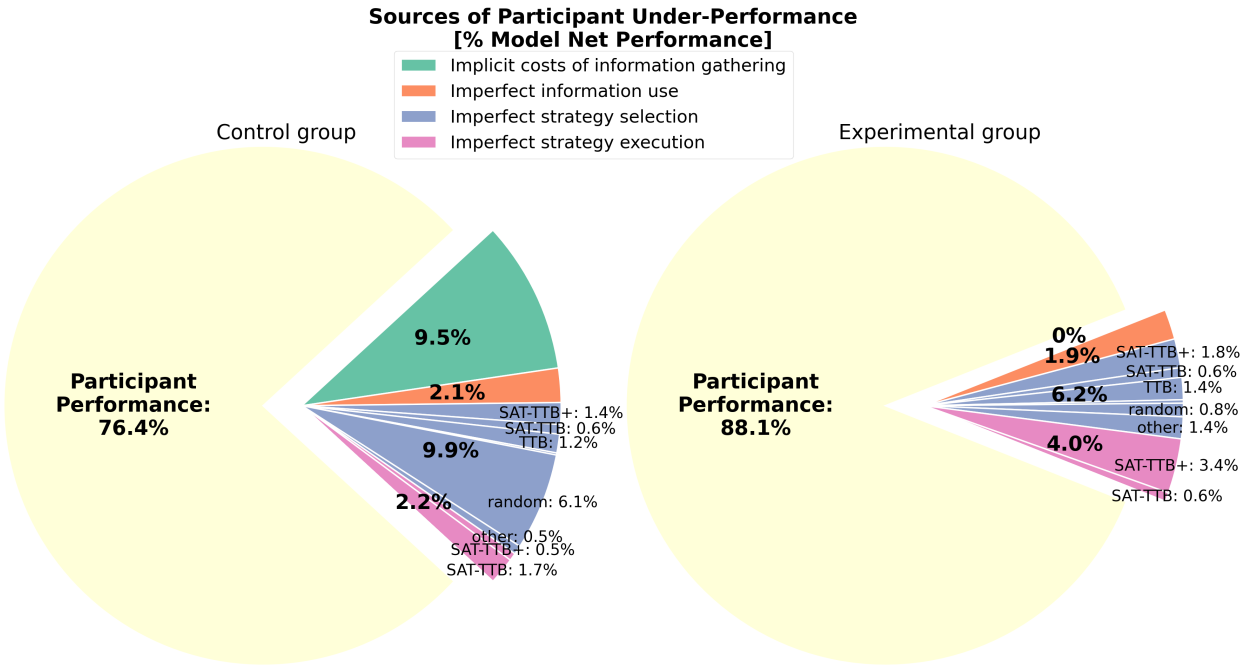

Figure E7

Same results as Figure 14 from Experiment 2, but excluding low-effort participants. 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.*

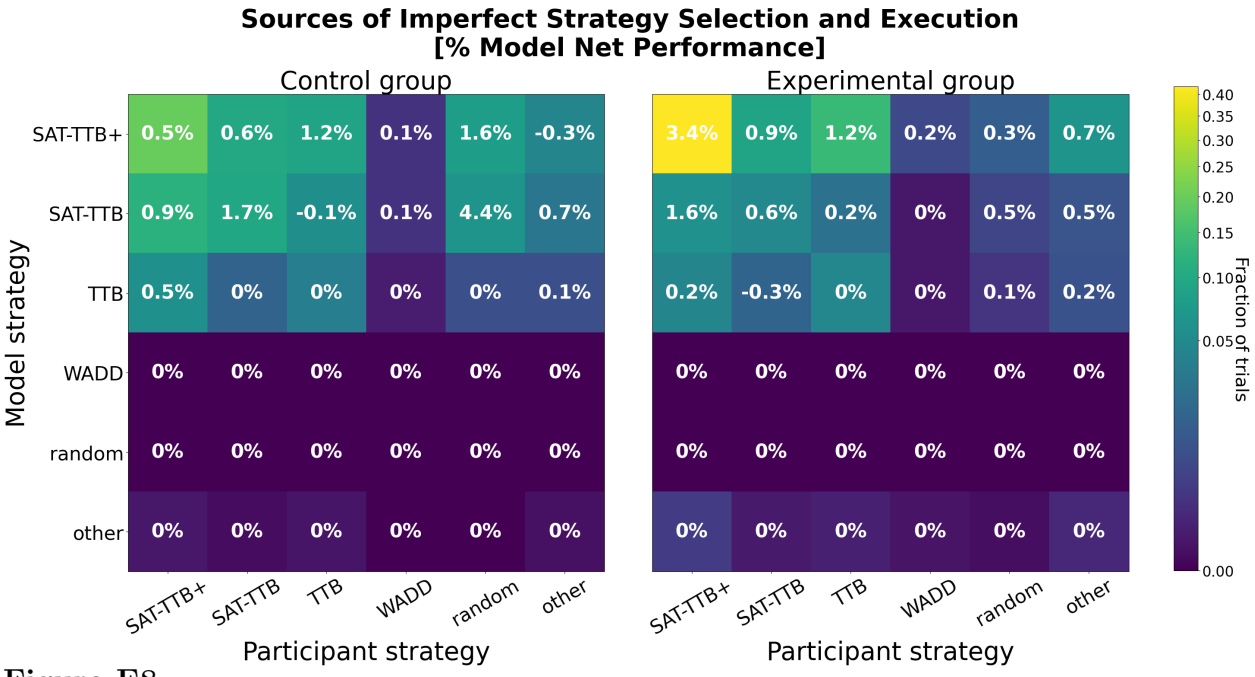

Figure E8

Same results as Figure 15 from Experiment 2, but excluding participants who did not perform the task correctly.