Discovering Rational Heuristics for Risky Choice

Paul M. Krueger[a,1,2], Frederick Callaway[b,1], Sayan Gul[c], Thomas L. Griffiths[a,b], and Falk Lieder[d]

[a]Department of Computer Science, Princeton University, Princeton, NJ, USA; [b]Department of Psychology, Princeton University, Princeton, NJ, USA; [c]Department of Psychology, University of California, Berkeley, CA, USA; [d]Max Planck Institute for Intelligent Systems, Tübingen, Germany

January 7, 2022

1 Abstract

For computationally limited agents such as humans, perfectly rational decision-making is almost always out of reach. 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 machine learning method that identifies the best heuristics to use in any given situation. To demonstrate the generalizability and accuracy of our method, we compare the strategies it discovers against those used by people across a wide range of multi-alternative 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 they tend to collect too little information and their strategy choices do not always fully exploit the structure of the environment.

2 Introduction

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 (1–6). According to classic economic theory, people should choose their actions so as to maximize the expected value of the consequences (7, 8), but computing those expected values for real world problems is a substantial task and humans face significant limitations in computational resources and time (9). 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 (10–14).

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 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 (9, 15–21) – a normative benchmark that we refer to as "resource rationality" (16, 21). 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 (22). Specifically, heuristic decision-making can itself be understood as a sequential decision problem (23). 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 method, we applied it to multi-alternative, multi-attribute decision-making (24). 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 (see Figure 1) (25). 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 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.

Previous work has manually identified a number of heuristics employed in multi-alternative risky choice (26–29), and characterized the environments in which hand-chosen heuristics perform best, showing that people select among those heuristics accordingly (27, 30–37). More recent work has sought to automate the characterization of known heuristics using a neural network algorithm (38). It still remains unknown, however, whether other effective heuristics remain undiscovered, and whether known heuristics correspond to a normative standard. Our method extends these previous results by discovering the best-performing heuristics from an immense, combinatorial strategy space defined by a set of basic cognitive operations. Applying reinforcement learning to arbitrary discrete steps of cognitive operations provides a formalism that can be applied to any setting where the goal is to characterize the optimality of heuristics, while circumventing the need to search a huge combinatorial space. This formalism also offers a normative standard for evaluating heuristics.

To automatically search for potentially undiscovered heuristics, we recently developed a reinforcement learning algorithm that approximates the value of information by interpolating between the myopic value of information and the value of perfect information (39). We applied this method across a very large range of scenarios and tested its predictions in an experiment of unprecedented scale, a full order of magnitude larger than the largest previous study in this setting. We collected data from over 2,300 participants, systematically varying the parameters of the decision-making environment. This allowed us to explore a large space of potential heuristics that may be employed in the Mouselab task. It further 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 method automatically rediscovered the classic Take-The-Best (TTB) and Weighted-Additive (WADD) heuristics (11) as resource-rational strategies in specific situations. In addition, our method discovered novel heuristics that had been previously overlooked. Importantly, 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 these strategies imperfectly and generally gather too little information, thus falling short of perfect resource-rational decision-making. These 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.

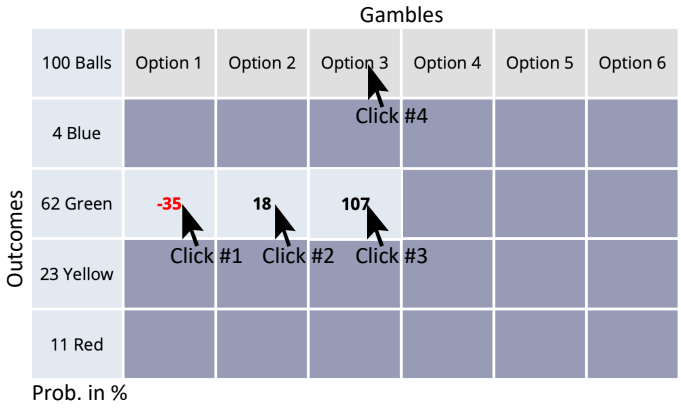

Fig. 1. Illustration of the Mouselab paradigm. The task is to choose one of six gambles, each of which results in one of four probabilistic outcomes; before gambling, participants can gather information about the value of each cell by clicking on it. The Mouselab paradigm externalizes computations by 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.

Automatic strategy discovery

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). 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 (40), 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" (41), 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 (42). 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., (43–46)) and success in describing human (e.g., (47, 48)) and animal (e.g., (49, 50)) behavior and brain function (e.g., (51–56)).

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). Methods from reinforcement learning that are used to solve MDPs can be built upon to solve meta-level

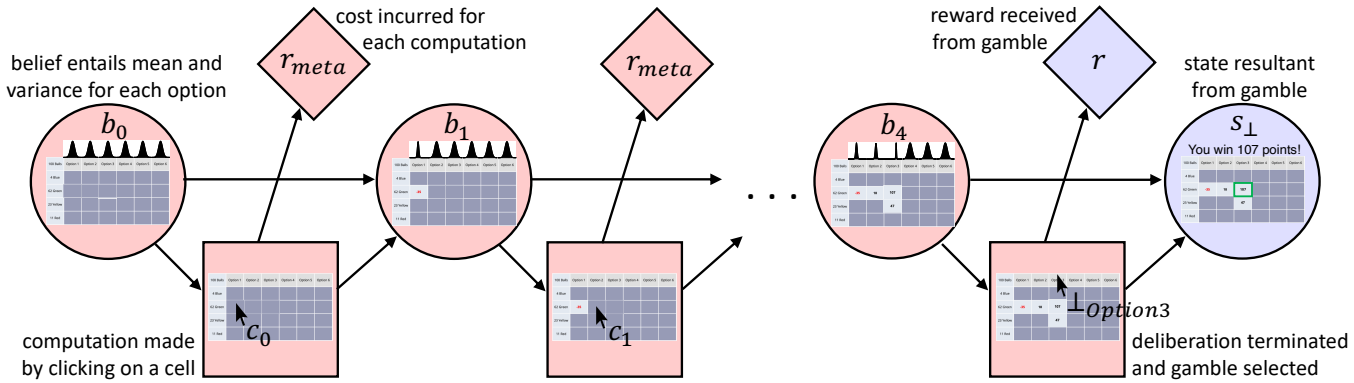

Fig. 2. Schematic illustration of the meta-level Markov Decision Process framework applied to the Mouselab task. At the beginning of each trial, when all cell values are hidden, the agent's initial belief state, b_0, is represented as Gaussian distribution for each of the six options. Each time the agent 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).

MDPs, thus providing a formal framework describing 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 (22, 40), 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 (57) as well as planning (58, 59). Here, we apply this approach to compute resource-rational heuristics for multi-attribute risky choice and compare them to the strategies that people use.

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) (39). 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.

Results

We set out to discover resource-rational heuristics by applying our computational strategy discovery method to the Mouselab task, and then clustering on the sequences of clicks generated by our method. We then compared those clusters of click sequences 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.

We designed a large-scale process-tracing experiment using the Mouselab task, with 6 options (or "alternatives") and 4 possible outcomes ("attributes") (Figure 1). To test how well human decision strategies correspond to the optimal heuristics derived by our method, we collected data across 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 cost, resulting in a total of fifty conditions (see Materials and Methods for details). For each condition, we applied our strategy discovery method by formulating a corresponding meta-level MDP and finding an approximately optimal solution using the BMPS algorithm (see Materials and Methods). We then presented 2,368 human participants with the same fifty conditions in a between-subjects design with about 47 participants per condition.

Identification of resource-rational decision strategies

Previous work has identified a set of well-known heuristics that people use in multi-alternative risky choice, including Take-the-Best (TTB, defined as choosing between alternative options based on the one single attribute that is the best predictor of the outcome* (28)) and Satisficing (SAT, defined as considering alternative options until it finds one that is good enough (29)), and other strategies, including Weighted Additive (WADD, which computes the expected payoffs of all alternatives (11, 29, 60)). 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

*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.

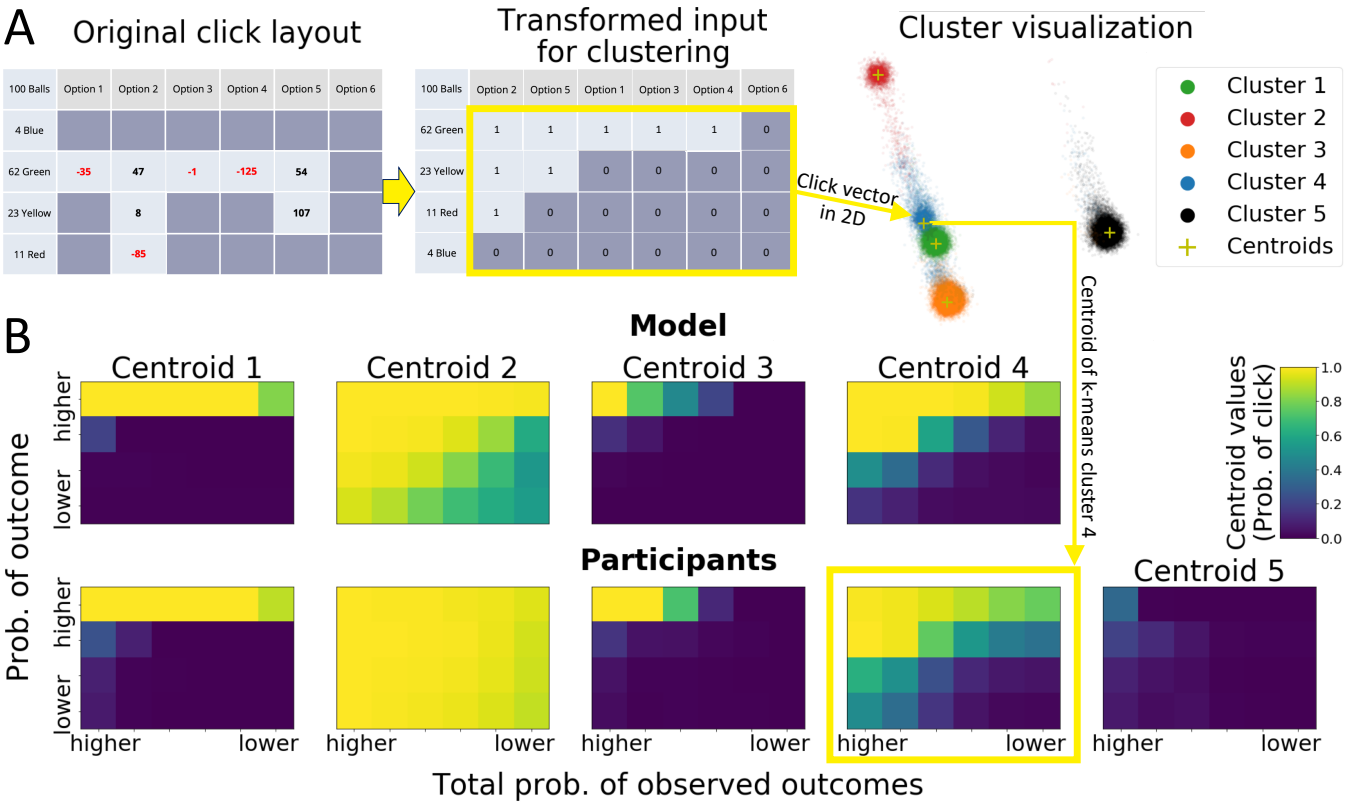

Fig. 3. Data-driven strategy identification. **(A)** The sequence of clicks on a given trial is converted into an indicator matrix with uninformative spatial variation removed. Rows are rearranged from the most to least probable outcome, and columns are rearranged in descending order of the sum of the 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 $\times 20$ trials per participant; visualized here projected onto 2D space via Fisher's Linear Discriminant Analysis) serve as input to a k-means clustering algorithm. **(B)** Centroids for the clusters uncovered in human data and model simulations. 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.

all fifty decision environments. To explore this space in a data-driven way, we applied the k-means clustering algorithm to the sequences of clicks performed by our resource-rational model and by human participants. 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.

Inspecting the heuristic prototypes from the resource-rational model revealed that our method rediscovered the TTB heuristic (11) and the WADD strategy. The sequences of clicks in each of these two clusters correspond closely to each of these two well known heuristics. Two additional prototypes revealed two previously undiscovered heuristics. The first, which we call SAT-TTB, combines elements of TTB and Satisficing (see Figure 1). Like TTB, SAT-TTB inspects only the payoffs for the most probable outcome. But unlike TTB and like Satisficing, SAT-TTB terminates as soon as it finds a gamble whose payoff for the most probable outcome is high enough, reducing the amount of information considered. 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).

Figure 3A illustrates the procedure used to transform click sequences into a 24-dimensional binary vector with invariant ordering of rows and columns, which comprised the inputs to k-means clustering. Figure 3B shows a 2-dimensional embedding of these vectors from every trial, highlighting which k-means cluster each vector belongs to. Figure 3C shows the centroids identified by applying k-means clustering to resource-rational click sequences (top) and human click sequences (bottom), revealing a close correspondence between the strategies deployed by the resource-rational model and participants' strategies. Centroid 1 corresponds to the TTB strategy, where participants inspect only the most probable attribute for each alternative option. Centroid 2 corresponds to the WADD strategy, which clicks practically everywhere, hence the nearly all-yellow color. Centroids 3 and 4 correspond to the two newly discovered strategies: 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. The two newly discovered heuristics do not correspond to any previously known heuristics. Yet, as described below, we found that people frequently use these heuristics across the wide range of environments in which they are adaptive.

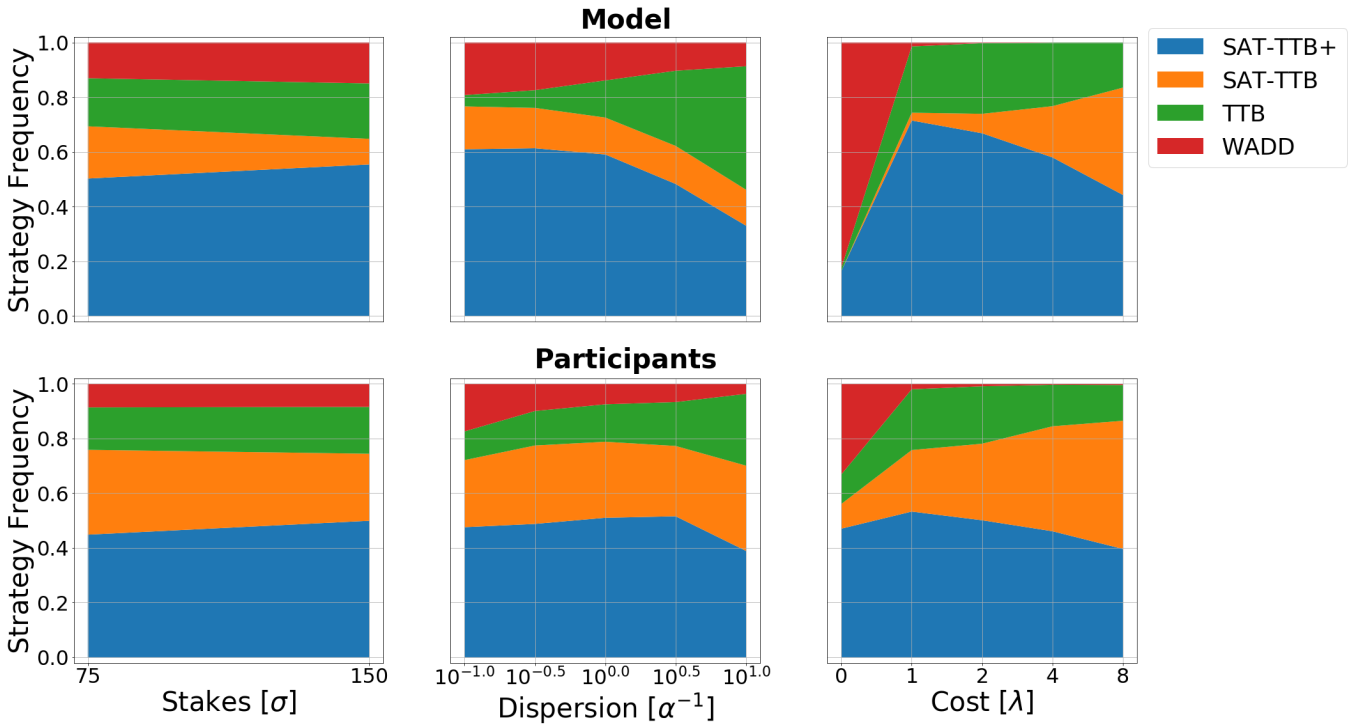

Fig. 4. Strategy use frequencies of the resource-rational model and human participants as a function of the three environment parameters: σ, the standard deviation of possible payoffs, α^{-1}, the peakiness of the outcome distribution, and $lambda$, the cost paid for each piece of information revealed.

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).[†]

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 the 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 (unstandardized slope $\beta = -0.0061, p < 0.001$) and a significant positive effect on the frequency of SAT-TTB+ ($\beta = 0.0035, p < 0.001$; left panels of Figure 4).

The model predicted that as the outcome distribution becomes more peaky (i.e., higher dispersion), the use of TTB should steadily increase; intuitively, one can focus on a single outcome when only one is likely to occur. Our participants confirmed this prediction ($\beta = 0.20, p < 0.001$; 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 ($\beta = -0.26, p < 0.001$).

When there is no cost for gathering information, the model usually uses WADD, very rarely using this strategy otherwise.[‡] 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+ ($\beta = -0.11, p < 0.001$, respectively) and TTB ($\beta = -0.051, p < 0.001$), while increasing use of the most frugal strategy, SAT-TTB ($\beta = 0.26, p < 0.001$). Figures S1 and S2 compare strategy frequencies in each of the 50 conditions, showing broad correspondence between the resource-rational model and participants.

Table S1 summarizes post-hoc pairwise comparisons and effect sizes for the statistics reported in this section.

Rational strategy selection explains variability in choice behavior

Previous research on multi-alternative risky choice has characterized people's choice behavior in the Mouselab paradigm in terms of four features (25, 61, 62). 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 alternatives and attributes, respectively. Our resource-rational model predicts how these behavioral characteristics should vary between different decision environments. To test these predictions, we assessed the contingency of these qualitative aspects of

[†] 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.

[‡] The resource-rational model does not always use WADD when the cost is zero because in some scenarios the expected value of information is less than floating point precision.

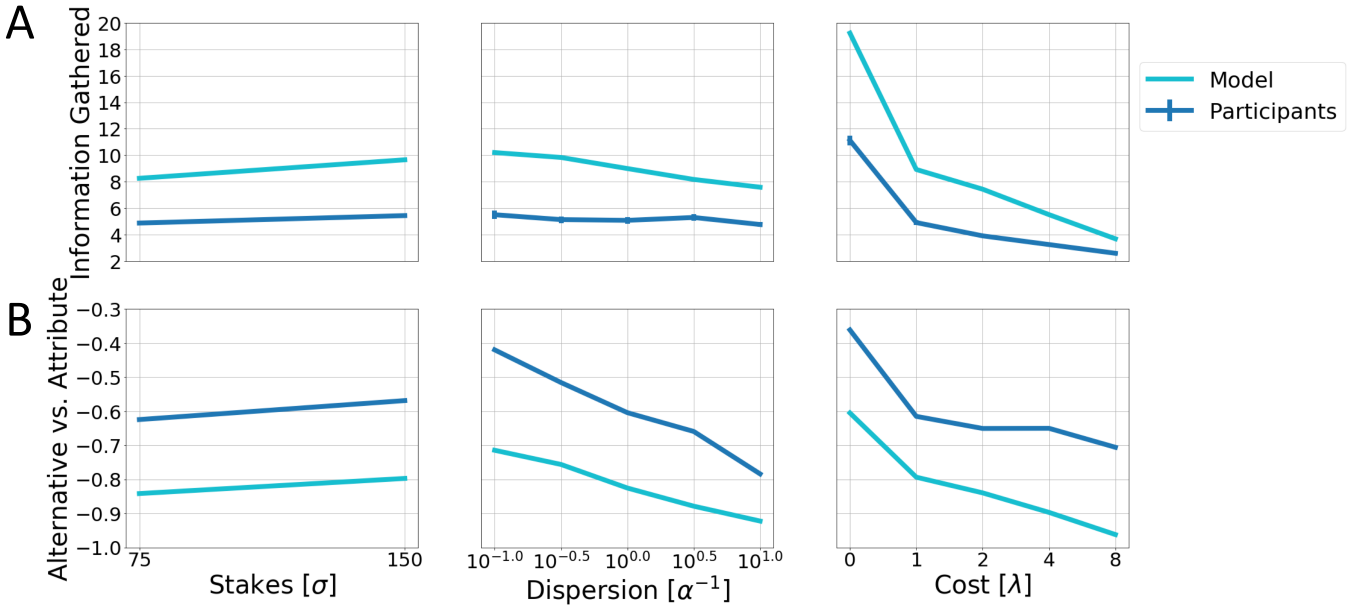

Fig. 5. Behavioral correspondence between participants and the resource-rational model. **(A)** The average number of values revealed by participants and the model as a function of each environment parameter. **(B)** The same, but for a measure of alternative- vs. attribute-based processing (negative indicates attribute-based). Error-bars show the SEM across participants.

people's decision-making style on the structure of the decision problem.

We first considered 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 regression with random intercepts for participants revealed that the stakes significantly predicted information gathered: standardized $\beta = 0.0013, p = 0.009$). 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 ($\beta = -0.008, p = 0.098$). Finally, people reduced information gathering as it became more costly to do so ($\beta = -0.13, p < 0.001$). However, across all conditions, participants made on average 3.80 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 options along one dimension. In alternative-based processing, sequential clicks are made on one column/gamble; this corresponds to evaluating one option 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 (25, 61). 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 ($\beta = 0.001, p = 0.016$); they used more attribute-based processing as dispersion increased ($\beta = -0.043, p < 0.001$) and the cost increased ($\beta = -0.047, p < 0.001$). A comparison of information gathering and alternative- vs. attribute-based processing for the model and participants across each of the fifty decision environments in shown in Figure S9, 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 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) (25, 61). Alternative variance 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 S10 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; $\beta = -0.002, p < 0.001$) and alternatives (alternative variance; $\beta = -0.0012, p = 0.0016$),

likely due to an overall increase in information gathering. When one outcome was much more likely than all others then 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 ($\beta = 0.041, p < 0.001$) and more evenly across the alternatives ($\beta = -0.018, p < 0.001$). Finally, increasing the cost of information made people more discerning in how much attention they paid to different attributes ($\beta = 0.096, p < 0.001$) and different alternatives ($\beta = 0.095, p < 0.001$). Figure S11 shows the qualitative correspondence between the model and participants for these two measures across all fifty decision environments.

Sources of under-performance

In addition to providing a framework for discovering heuristics, our formalism provides a normative standard to compare with participants' performance. On average, participants achieved 57.1% as many points as the resource-rational agent (see Figure S12 for a comparison across conditions). What explains this sizable gap? As detailed in the SI and summarized in Table S3, we identified four possible causes of participant under-performance: implicit costs, imperfect information use, imperfect strategy selection, and imperfect strategy execution. We briefly summarize these results below.

First, participants may be influenced by costs not accounted for by our resource-rational model, which could explain why participants collected less information than the resource-rational agent. These costs could include, for example, physical constraints like the effort required to move a cursor and make clicks, as well as cognitive costs associated with processing the revealed information (25). As a first step towards quantifying the extent to which such costs hindered our participants' performance, we simulated the model with an implicit cost-per-click set to match the average amount of information people gathered. This model had a 9.0% reduction in performance, which accounts for 23.0% of the performance gap between participants and the model. Importantly, this number approximates the proportion of the performance gap attributable to implicit costs under the assumption that people act optimally with respects to those costs; 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.

A second source of under-performance is 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. However, this source accounted for only 6.8% of participants' reduced performance.

A third possible source of suboptimality 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. Indeed, we find that imperfect strategy selection accounts for 63.0% of the performance differential between participants and the model. The majority of this gap (59.0%, or 37.2% of the total performance gap) is attributed to trials in which participants gamble randomly when they should have considered some information.[§]

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 accounts for the remaining 7.2% of under-performance. Nearly all of this gap is attributed to imperfect execution of the SAT-TTB+ strategy, which is the most complex strategy to execute.

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 make a choice 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 from a rational analysis of how agents should allocate 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 rediscover existing heuristics and identify new heuristics that had previously been overlooked. Furthermore, we collected a dataset of unprecedented scale 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 follow all of the same heuristics as our method, and adaptively select which heuristic to use in a way that is consistent with our framework, but that 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 a meta-level Markov Decision Process (22, 40). 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 has previously been studied from a resource-rational perspective (63), the

[§]The SI also presents results that exclude participants who gamble randomly on more than half of all trials.

meta-MDP framework provides a new set of computational tools for understanding heuristics through this lens. The result is that we are able to formally identify heuristics that achieve an optimal trade-off between computational costs and decision quality. Being able to derive heuristics directly from a normative model forms a direct contrast to the cumbersome and inexact process of searching for heuristics by hand that psychologists have relied on in the past.

We demonstrated the usefulness of this approach using the Mouselab task, which is a classic, well-studied process tracing paradigm (64). While the Mouselab task has been widely used to study decision strategies, these studies are typically limited to around $20-40$ participants (e.g. (35, 60, 62, 65–68)), rarely exceed 100 (69–71), 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 (72). 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.

In the present task, participants used the same four strategies as the resource-rational model, and it is useful to consider their source. It is typically assumed that people have a limited toolkit of general-purpose heuristics that are adapted to real-world environments (e.g., (73, 74)). 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 (75). 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 S12 and S13). This suggests that human decision-making still has room for improvement. 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 (76–79).

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. Nevertheless, it is worth considering the specific sources for this gap, which are detailed in the SI section *Sources of under-performance*. 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 (80), 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 (81, 82). 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.

Our resource-rational framework both offers a normative standard for evaluating heuristics and, importantly, rests on a formalism that makes it generally applicable to any decision-making process. Researchers have previously considered the ideal observer perspective for rational decision-makers (83–85), but such an approach was recognized as infeasible (86–89). An alternative view is to emphasize the limitations of the decision-maker and the fact that heuristics are computationally cheaper (60, 64) and may achieve some trade-off between accuracy and effort (90, 91) or optimization under constraints due to information costs (92, 93), although these perspectives typically view heuristics as inferior to rational decisions (94, 95). The discovery that simpler regression models may outperform more complex ones (96–99), combined with observations that heuristics often work quite well in many real-world decision environments (100–106)—the so-called "less-is-more" effect—challenged the classical normative view of rationality. This led to the idea of ecological rationality (64, 107, 108), and attempts to account for the effectiveness of heuristics in terms of the structure of the decision environment (27, 30–35), the effectiveness of reducing model parameters to balance the bias-variance trade-off (36, 109) or when observations are limited or noisy (110–113), and Bayesian inference with strong priors (114). More recently, a resource-rational analysis of cognition has been applied to view heuristics as making rational use of limited computational resources (21, 38, 115, 116). The formalism used here for our resource-rational approach, defined by the meta-MDP, 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 formalism for deriving optimal heuristics that avoids the need to search an intractable combinatorial space of possible heuristics, as well as a normative benchmark for evaluating heuristics.

The finding that participants use resource-rational heuristics in an adaptive manner suggests that people have highly effective mechanisms for discovering and selecting good heuris-

tics. 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.

Materials and methods

Strategy discovery

We have developed an automatic method for discovering rational decision strategies. We define rational decision strategies as the solution to the problem of deciding how to decide. Given certain assumptions about the decision-maker's cognitive architecture and the structure of the decision environment, we can we formalize this problem as a meta-level Markov Decision Process (MDP). To compute the resource-rational heuristic entailed by this definition, we then apply the Bayesian meta-level policy search algorithm (39) to approximate the optimal policy for selecting the next decision operation given the decision-maker's current knowledge state.

Here, we applied this approach to the Mouselab paradigm (25). The following paragraphs explain how we modelled the problem of meta-decision-making in the Mouselab paradigm as a meta-level MDP, how we solved this problem using BMPS, and how we characterized the resulting solution in terms of simple decision strategies.

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 if the event corresponding to its row was to occur. Critically, all of 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.

Meta-level MDP model of multi-attribute risky choice in the Mouselab paradigm Before defining our meta-level MDP model, we briefly review generic Markov Decision Processes (MDPs) (117). MDPs are the standard formalism for modeling sequential decision problems, in which an agent iteratively interacts with an environment in order to maximize attained reward. An MDP is defined by a four-tuple, $M_{object} = \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}$, where S is a set of possible environment states, $\mathcal{A}$ is a set of actions that an agent can take, $\mathcal{T}$ 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 $\mathcal{R}$ is a reward function describing the reward received for such a transition: $\mathcal{R}(s, a)$. 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. (The term "object-level MDP" unambiguously refers to a generic MDP that is not a meta-level MDP (41)). A meta-level MDP is defined by a four-tuple, $M_{\text{meta}} = B, C, T_{\text{meta}}, r_{\text{meta}}$. Here, states are replaced by a set of beliefs, B, describing what the agent may think; actions are replaced by a set of computations, C, describing cognitive operations the agent can perform; the meta-level transition function, T_{meta}, describes how a computation, c, made with belief b leads 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 action that has highest expected utility given the current belief; thus $r_{\text{meta}}(b_t, \perp) = \max_a \mathbb{E}_{s \sim b_t}[U(s, a)]$ 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, $r_{\text{meta}}(b_t, \perp)$, 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 ($\mathcal{O}$) and probability distributions $(b_{t,1}, \cdots, b_{t,n})$ over the expected values $e_1 = \mathbb{E}[v_{O,g_1}], \cdots e_n = \mathbb{E}[v_{O,g_n}]$ of the n available gambles $g_1, \cdots, g_n$, where v_{O,g_i} contains the values that have been observed for gamble $i \in 1 : n$. Furthermore, we assume that for each element $v_{o,g}$ of the payoff matrix V 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 normal 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

$$b_{t,g} = \left(b_{t,g}^{(\mu)}, b_{t,g}^{(\sigma^2)}\right), \qquad [1]$$

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

$$b_{t,g}^{(\mu)} = \sum_{(o,g)\in\mathcal{O}} p(o)\cdot v_{o,g} + \sum_{(o,g)\notin\mathcal{O}} p(o)\cdot \bar{v} \quad [2]$$

$$b_{t,g}^{(\sigma^2)} = \sum_{(o,g)\notin\mathcal{O}} p(o)^2\cdot \sigma_v^2. \quad [3]$$

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 $c_{o,g} \in C$ 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 resource-rational heuristics $h^\star$ as optimal meta-level policies that maximize the meta-level reward for making a decision in an well-informed belief state minus the cost of attaining it, that is

$$\begin{aligned} h^\star &= \arg\max_{\pi_{(\text{meta})}} \mathbb{E}\left[\sum_t r_{(\text{meta})}(b_t, \pi_{(\text{meta})}(b_t))\right] \\ &= \arg\max_{\pi_{(\text{meta})}} \mathbb{E}\left[\max_g b_T^{(\mu)}(g) - T\cdot\lambda\right], \end{aligned} \quad (4)$$

where the random variable T 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 (39).

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 (39). 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, BMPS optimizes the meta-level return of the meta-level policy

$$\begin{aligned} \pi_{\text{meta}}(b) = \arg\max_c\, & w_1\cdot \text{VOC}_1(b,c) + w_2 \\ & \cdot \text{VPI}_{sub}(b,c) + w_3\cdot \text{VPI}(b) - w_4\cdot \text{cost}(c), \end{aligned} \quad (5)$$

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 determines the weights $w_1, \cdots, w_4$ by maximizing the expected meta-level return of the resulting meta-level policy using Bayesian optimization.

Application of BMPS to the Mouselab paradigm To compute optimal risky choice strategies we applied BMPS to a meta-level MDP model of decision-making in the Mouselab paradigm described above. 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 gambles, rather than deciding based on the current belief state. Formally, it is

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

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

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

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

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

$$\text{VPI}_{\text{sub}}(b_t, c) = \mathop{\mathbb{E}}_{v_{g_c}^*|b_{t,g_c}}\left[\max\left\{v_{g_c}^*, \max_{g\neq g_c} b_{t,g}^{(\mu)}\right\}\right] - \max_g b_{t,g}^{(\mu)}, \quad [6]$$

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

Finally, the cost of computation feature was simply

$$\text{cost}(c_t) = r_{\text{meta}}(b, c_t) = -\lambda.$$

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. The algorithm then returns the weight vector with highest expected performance. See (39) for details of the BMPS optimization procedure.

To derive the optimal heuristics for each environment we then characterized the click-behavior of the best-performing meta-level policy that we found across all runs of BMPS.

Clustering and definition of strategies To identify distinct strategies, we applied the Elkan k-means clustering algorithm to the locations of clicks predicted by our resource-rational model across trials, with an Euclidean distance metric (118). The following steps were performed to reduce uninformative spatial variation across trials in the locations of clicks. First, for each trial 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 least probable outcome, and columns (gambles) were rearranged in descending order of the sum of the probabilities of the outcomes observed in that column (this transformation is illustrated in Figure 3A). 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. Fisher's Linear Discriminant Analysis (LDA) was used to project the 24-dimensional click sequence vectors onto a 2-dimensional space (119). For the resource-rational model, we selected $k = 4$ clusters, because this identified unique types of click patterns; $k > 4$ resulted in redundant patterns of clicking, which could be due to a limit in the number of strategies people use, or a limitation of the clustering method. Similarly, for human participants using $k = 5$ clusters produced distinct click patterns whereas using $k > 5$ clusters resulted in groups of redundant strategies. For visualization purposes, the centroid of each cluster was reshaped into a 4×6 matrix (Figure 3C).

Based on the clustering solutions, we defined 5 distinct strategies as follow: 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, then selecting the gamble with the highest observed value; 4) WADD was operationalized as selecting nearly all cells; specifically, the strategy was classified as WADD for a given sequence of clicks if BMPS deemed that additional clicks would yield negligibly small value (i.e., when the value of additional clicks was less than \$0.01; the average such threshold was 19.1 clicks); 5) A random strategy entailed zero clicks, and 6) other strategies were those not consistent with any of the previous five definitions.

Behavioral experiment

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).

Stimuli and procedure Following instructions and a comprehension check, participants performed a variation of the Mouselab task (25). 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. 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 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 subtracting for each click ($\lambda \in \{0, 1, 2, 4, 8\}$). This created a total of $2 \times 5 \times 5 = 50$ conditions.

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, and how some outcomes were more likely than others. Participants were given three practice trials, and after these instructions were given a quiz that assessed their understanding of all critical information conveyed in the instructions. The full experiment, including instructions, can be viewed here: https://mouselab.herokuapp.com/. 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.

Acknowledgements 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.

References

[1] Peter Bossaerts and Carsten Murawski. Computational complexity and human decision-making. *Trends in Cognitive Sciences*, 21(12):917–929, 2017.

[2] Peter Bossaerts, Nitin Yadav, and Carsten Murawski. Uncertainty and computational complexity. *Philosophical Transactions of the Royal Society B*, 374(1766):20180138, 2019.

[3] Sebastian Nowozin. Optimal decisions from probabilistic models: the intersection-over-union case. In *Proceedings of the*

IEEE conference on computer vision and pattern recognition, pages 548–555, 2014.

[4] Johan Kwisthout, Todd Wareham, and Iris van Rooij. Bayesian intractability is not an ailment that approximation can cure. *Cogn. Sci.*, 35(5):779–784, 2011.

[5] Christos H Papadimitriou and John Tsitsiklis. Intractable problems in control theory. *SIAM journal on control and optimization*, 24(4):639–654, 1986.

[6] Samuel J Gershman, Eric J Horvitz, and Joshua B Tenenbaum. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349 (6245):273–278, 2015.

[7] Oskar Morgenstern and John Von Neumann. *Theory of games and economic behavior*. Princeton university press, 1953.

[8] Leonard J Savage. The theory of statistical decision. *Journal of the American Statistical association*, 46(253):55–67, 1951.

[9] Herbert A Simon. Theories of bounded rationality. *Decision and organization*, 1(1):161–176, 1972.

[10] Thomas Gilovich, Dale Griffin, and Daniel Kahneman. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press, 2002.

[11] Gerd Gigerenzer and Daniel G Goldstein. Betting on one good reason: The take the best heuristic. In *Simple heuristics that make us smart*, pages 75–95. Oxford University Press, 1999.

[12] AJ Maule and GP Hodgkinson. Heuristics, biases and strategic decision making. *Psychologist*, 15(2):68–71, 2002.

[13] Justin L Gardner. Optimality and heuristics in perceptual neuroscience. *Nature neuroscience*, 22(4):514–523, 2019.

[14] Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982.

[15] Herbert A. Simon. Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138, 1956. .

[16] Thomas L Griffiths, Falk Lieder, and Noah D Goodman. Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, 7(2):217–229, 2015.

[17] Thomas L Griffiths, Edward Vul, and Adam N Sanborn. Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science*, 21(4):263–268, 2012.

[18] Carlos Zednik and Frank Jäkel. Bayesian reverse-engineering considered as a research strategy for cognitive science. *Synthese*, 193(12):3951–3985, 2016.

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

[20] Richard L Lewis, Andrew Howes, and Satinder Singh. Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in cognitive science*, 6 (2):279–311, 2014.

[21] Falk Lieder and Thomas L Griffiths. Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43, 2020.

[22] Stuart Russell and Eric Wefald. Principles of metareasoning. *Artificial intelligence*, 49(1-3):361–395, 1991.

[23] Thomas L Griffiths, Frederick Callaway, Michael B Chang, Erin Grant, Paul M Krueger, and Falk Lieder. Doing more with less: meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences*, 29:24–30, 2019.

[24] Stelios H Zanakis, Anthony Solomon, Nicole Wishart, and Sandipa Dublish. Multi-attribute decision making: A simulation comparison of select methods. *European journal of operational research*, 107(3):507–529, 1998.

[25] John W. Payne, James R. Bettman, and Eric J. Johnson. Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):534–552, 1988. .

[26] John W Payne. Heuristic search processes in decision making. *ACR North American Advances*, 1976.

[27] Konstantinos V Katsikopoulos. Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice. *Decision analysis*, 8(1):10–29, 2011.

[28] Gerd Gigerenzer and Daniel G Goldstein. Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4):650–669, 1996.

[29] Herbert A Simon. Rational choice and the structure of the environment. *Psychological review*, 63(2):129–138, 1956.

[30] Laura Martignon and Ulrich Hoffrage. Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52(1):29–71, 2002.

[31] Laura Martignon, Ulrich Hoffrage, ABC Research Group, et al. Why does one-reason decision making work. *Simple heuristics that make us smart*, pages 119–140, 1999.

[32] Konstantinos V Katsikopoulos and Laura Martignon. Naive heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, 50(5): 488–494, 2006.

[33] Manel Baucells, Juan A Carrasco, and Robin M Hogarth. Cumulative dominance and heuristic performance in binary multiattribute choice. *Operations research*, 56(5):1289–1304, 2008.

[34] Özgür Şimşek. Linear decision rule as aspiration for simple decision heuristics. *Advances in neural information processing systems*, 26:2904–2912, 2013.

[35] Anja Dieckmann and Jörg Rieskamp. The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, 35(7):1801–1813, 2007.

[36] Gerd Gigerenzer and Henry Brighton. Homo heuristicus: Why biased minds make better inferences. *Topics in cognitive science*, 1(1):107–143, 2009.

[37] Daniel G Goldstein and Gerd Gigerenzer. Models of ecological rationality: the recognition heuristic. *Psychological review*, 109(1):75–90, 2002.

[38] Eric Schulz Marcel Binz, Samuel J. Gershman and Dominik Endres. Heuristics from bounded meta-learned inference. *PsyArXiv*, 2020, August 6.

[39] Frederick Callaway, Sayan Gul, Paul M. Krueger, Thomas L. Griffiths, and Falk Lieder. Learning to select computations. In *Uncertainty in Artificial Intelligence*. UAI Press, 2018.

[40] Nicholas Hay, Stuart Russell, David Tolpin, and Solomon Shimony. Selecting computations: Theory and applications. In Nando de Freitas and Kevin Murphy, editors, *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*. AUAI Press, Corvallis, OR, 2012.

[41] Stuart Jonathan Russell and Eric Wefald. *Do the right thing: studies in limited rationality*. MIT press, 1991.

[42] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[43] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676): 354–359, 2017.

[44] Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemysław Dębiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, et al. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*, 2019.

[45] Matteo Hessel, Joseph Modayil, Hado Van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. Rainbow: Combining improvements in deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[46] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin

Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[47] Michael X Cohen and Charan Ranganath. Reinforcement learning signals predict future decisions. *Journal of Neuroscience*, 27(2):371–378, 2007.

[48] Hanan Shteingart and Yonatan Loewenstein. Reinforcement learning and human behavior. *Current Opinion in Neurobiology*, 25:93–98, 2014.

[49] Robert A Rescorla. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, pages 64–99, 1972.

[50] Richard S Sutton and Andrew G Barto. Time-derivative models of Pavlovian reinforcement. In M. Gabriel and J. Moore, editors, *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, chapter 12, pages 497—-537. MIT Press, Cambridge, 1990.

[51] Peter Dayan and Nathaniel D. Daw. Decision theory, reinforcement learning, and the brain. *Cognitive, Affective and Behavioral Neuroscience*, 8(4):429–453, 2008.

[52] Yael Niv. Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154, 2009.

[53] Paul W Glimcher. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, 108 (Supplement 3):15647–15654, 2011.

[54] Matthew M Botvinick, Yael Niv, and Andew G Barto. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition*, 113(3):262–280, 2009.

[55] Elliot A Ludvig, Marc G Bellemare, and Keir G Pearson. A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. *Computational neuroscience for advancing artificial intelligence: Models, methods and applications*, pages 111–144, 2011.

[56] Wolfram Schultz, Peter Dayan, and P Read Montague. A neural substrate of prediction and reward. *Science*, 275(5306): 1593–1599, 1997.

[57] Frederick Callaway, Antonio Rangel, and Thomas L. Griffiths. Fixation patterns in simple choice reflect optimal information sampling. *PLOS Computational Biology*, 17(3):e1008863, March 2021. ISSN 1553-7358.

[58] F. Callaway, F. Lieder, P. Das, S. Gul, P. M. Krueger, and T. L. Griffiths. A resource-rational analysis of human planning. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, Austin, TX, 2018. Cognitive Science Society.

[59] Frederick Callaway, Bas van Opheusden, Sayan Gul, Priyam Das, Paul Krueger, Falk Lieder, and Thomas L Griffiths. Human planning as optimal information seeking. *PsyArXiv*, 2021, January 15.

[60] John W Payne, James R Bettman, and Eric J Johnson. Adaptive strategy selection in decision making. *Journal of experimental psychology: Learning, Memory, and Cognition*, 14(3): 534–552, 1988.

[61] John W Payne. Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational behavior and human performance*, 16(2):366–387, 1976.

[62] Gerald L Lohse and Eric J Johnson. A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, 68(1):28–43, 1996.

[63] Xavier Gabaix, David Laibson, Guillermo Moloche, and Stephen Weinberg. Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068, 2006.

[64] John W Payne, John William Payne, James R Bettman, and Eric J Johnson. *The adaptive decision maker*. Cambridge university press, 1993.

[65] Maik Bieleke, David Dohmen, and Peter M Gollwitzer. Effects of social value orientation (svo) and decision mode on controlled information acquisition—a mouselab perspective. *Journal of Experimental Social Psychology*, 86:103896, 2020.

[66] Nils Reisen, Ulrich Hoffrage, and Fred W Mast. Identifying decision strategies in a consumer choice situation. *Judgment and decision making*, 3(8):641–658, 2008.

[67] Amos Arieli, Yaniv Ben-Ami, and Ariel Rubinstein. Tracking decision makers under uncertainty. *American Economic Journal: Microeconomics*, 3(4):68–76, 2011.

[68] Jörg Rieskamp and Philipp E Otto. Ssl: a theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135(2):207–236, 2006.

[69] Ravi Dhar, Stephen M Nowlis, and Steven J Sherman. Comparison effects on preference construction. *Journal of consumer research*, 26(3):293–306, 1999.

[70] Sankar Sen. The effects of brand name suggestiveness and decision goal on the development of brand knowledge. *Journal of Consumer Psychology*, 8(4):431–455, 1999.

[71] Rui Mata, Lael J Schooler, and Jörg Rieskamp. The aging decision maker: cognitive aging and the adaptive selection of decision strategies. *Psychology and aging*, 22(4):796–810, 2007.

[72] Dan N Stone and Kathryn Kadous. The joint effects of task-related negative affect and task difficulty in multiattribute choice. *Organizational behavior and human decision processes*, 70(2):159–174, 1997.

[73] Gary Klein. Naturalistic decision making. *Human factors*, 50 (3):456–460, 2008.

[74] John MC Hutchinson and Gerd Gigerenzer. Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural processes*, 69(2):97–124, 2005.

[75] Gerd Gigerenzer and Reinhard Selten. *Bounded rationality: The adaptive toolbox*. MIT press, 2002.

[76] F. Lieder, F. Callaway, Yash Raj Jain, Paul M. Krueger, P. Das, S. Gul, and T.L. Griffiths. A cognitive tutor for helping people overcome present bias. In *The Fourth Multidisciplinary Conference on Reinforcement Learning and Decision Making.*, 2019.

[77] F. Lieder, F. Callaway, Yash Raj Jain, P. Das, G. Iwama, S. Gul, Paul M. Krueger, and T. L. Griffiths. Leveraging artificial intelligence to improve people's planning strategies, 2020. Manuscript in revision.

[78] Julian Skirzyński, Frederic Becker, and Falk Lieder. Automatic discovery of interpretable planning strategies. *Machine Learning*, pages 2641—-2683, 2021.

[79] Anirudha Kemtur, Yash Raj Jain, Aashay Mehta, Frederick Callaway, Saksham Consul, Jugoslav Stojcheski, and Falk Lieder. Leveraging machine learning to automatically derive robust planning strategies from biased models of the environment. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, 2020.

[80] Stuart Russell. Rationality and Intelligence : A Brief Update. In Müller V. C., editor, *Fundamental Issues of Artificial Intelligence*, pages 1–21. 2016.

[81] David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman, 1982.

[82] John R Anderson. *The adaptive character of thought*. Psychology Press, 2013.

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

[84] Peter C Fishburn. Foundations of decision analysis: along the way. *Management science*, 35(4):387–405, 1989.

[85] Wilson S Geisler. Sequential ideal-observer analysis of visual discriminations. *Psychological review*, 96(2):267–314, 1989.

[86] Herbert A Simon. Invariants of human behavior. *Annual review of psychology*, 41(1):1–20, 1990.

[87] David E Bell, Howard Raiffa, and Amos Tversky. *Decision making: Descriptive, normative, and prescriptive interactions*.

cambridge university Press, 1988.

[88] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

[89] George E Kimball. A critique of operations research. *Journal of the Washington Academy of Sciences*, 48(2):33–37, 1958.

[90] Anuj K Shah and Daniel M Oppenheimer. Heuristics made easy: an effort-reduction framework. *Psychological bulletin*, 134(2):207–222, 2008.

[91] Lee Roy Beach and Terence R Mitchell. A contingency model for the selection of decision strategies. *Academy of management review*, 3(3):439–449, 1978.

[92] George J Stigler. The economics of information. *Journal of political economy*, 69(3):213–225, 1961.

[93] John R Anderson. The adaptive nature of human categorization. *Psychological review*, 98(3):409–429, 1991.

[94] Ralph L Keeney, Howard Raiffa, and Richard F Meyer. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press, 1993.

[95] Amos Tversky. Elimination by aspects: A theory of choice. *Psychological review*, 79(4):281–299, 1972.

[96] Robyn M Dawes and Bernard Corrigan. Linear models in decision making. *Psychological bulletin*, 81(2):95–106, 1974.

[97] Robyn M Dawes. The robust beauty of improper linear models in decision making. *American psychologist*, 34(7):571–582, 1979.

[98] Hillel J Einhorn and Robin M Hogarth. Unit weighting schemes for decision making. *Organizational behavior and human performance*, 13(2):171–192, 1975.

[99] Frank L Schmidt. The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, 31(3):699–714, 1971.

[100] Jean Czerlinski, Gerd Gigerenzer, and Daniel G Goldstein. How good are simple heuristics? In *Simple heuristics that make us smart*, pages 97–118. Oxford University Press, 1999.

[101] Jan M Lichtenberg and Özgür Şimşek. Simple regression models. In *Imperfect decision makers: Admitting real-world rationality*, pages 13–25. PMLR, 2017.

[102] Nick Chater, Mike Oaksford, Ramin Nakisa, and Martin Redington. Fast, frugal, and rational: How rational norms explain behavior. *Organizational behavior and human decision processes*, 90(1):63–86, 2003.

[103] Gerd Gigerenzer. *Rationality for mortals: How people cope with uncertainty*. Oxford University Press, 2008.

[104] Markus Wübben and Florian v Wangenheim. Instant customer base analysis: Managerial heuristics often "get it right". *Journal of Marketing*, 72(3):82–93, 2008.

[105] Michael D Lee*, Natasha Loughlin, and Ingrid B Lundberg. Applying one reason decision-making: the prioritisation of literature searches. *Australian Journal of Psychology*, 54(3): 137–143, 2002.

[106] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5): 1915–1953, 2009.

[107] Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic decision making. *Annual review of psychology*, 62:451–482, 2011.

[108] Gerd Gigerenzer and Peter M Todd. *Simple heuristics that make us smart*. Oxford University Press, USA, 1999.

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

[110] Robin M Hogarth and Natalia Karelaia. Ignoring information in binary choice with continuous variables: When is less "more"? *Journal of Mathematical Psychology*, 49(2):115–124, 2005.

[111] Robin M Hogarth and Natalia Karelaia. "take-the-best" and other simple strategies: Why and when they work "well" with binary cues. *Theory and Decision*, 61(3):205–249, 2006.

[112] Robin M Hogarth and Natalia Karelaia. Heuristic and linear models of judgment: Matching rules and environments. *Psychological review*, 114(3):733–758, 2007.

[113] Özgür Şimşek and Marcus Buckmann. Learning from small samples: An analysis of simple decision heuristics. *Advances in neural information processing systems*, 28:3159–3167, 2015.

[114] Paula Parpart, Matt Jones, and Bradley C Love. Heuristics as bayesian inference under extreme priors. *Cognitive psychology*, 102:127–144, 2018.

[115] F. Lieder and T. L. Griffiths. Strategy selection as rational metareasoning. *Psychological Review*, 124(6):762–794, 2017.

[116] Rahul Bhui, Lucy Lai, and Samuel J Gershman. Resource-rational decision making. *Current Opinion in Behavioral Sciences*, 41:15–21, 2021.

[117] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

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

[119] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

Supporting Information

Krueger, Callaway, Gul, Griffiths & Lieder

January 7, 2022

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 a 2-dimensional embedding of the clustering, for visualization purposes, for participants (Figures S1 and S2) and the resource-rational model (Figures S3 and S4), with each point representing the click sequence vector from a single trial, with the color of each point corresponding to its k-means cluster number (Figures S1 and S3) or strategy classification (Figures S2 and S4).

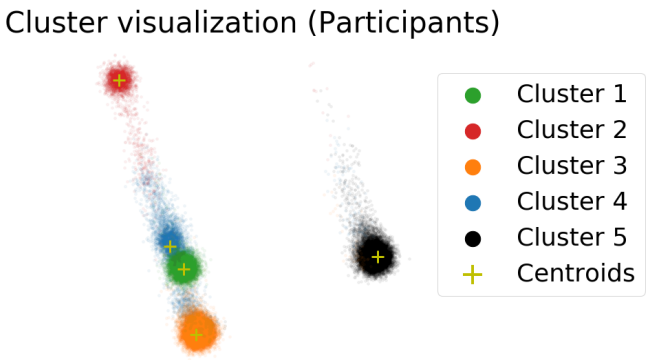

Fig. S1. Click sequence vectors from participant trials projected onto the 2D LDA transform, with labels corresponding to the k-means cluster number (as shown in Figure 3A).

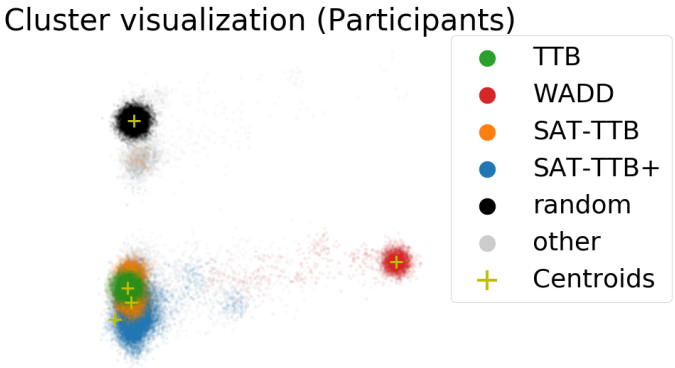

Fig. S2. Click sequence vectors from participant trials projected onto the 2D LDA transform, with labels corresponding to the strategy definition.

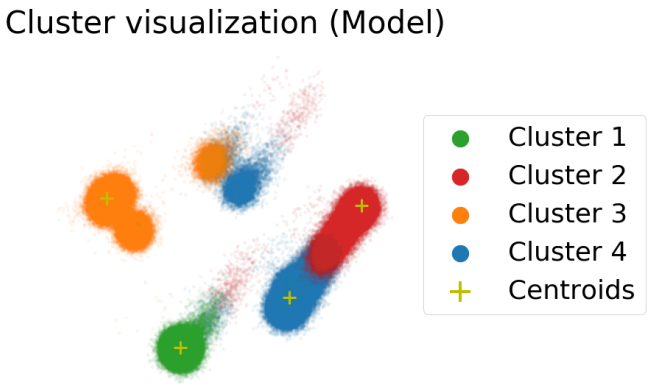

Fig. S3. Click sequence vectors from model trials projected onto the 2D LDA transform, with labels corresponding to the k-means cluster number.

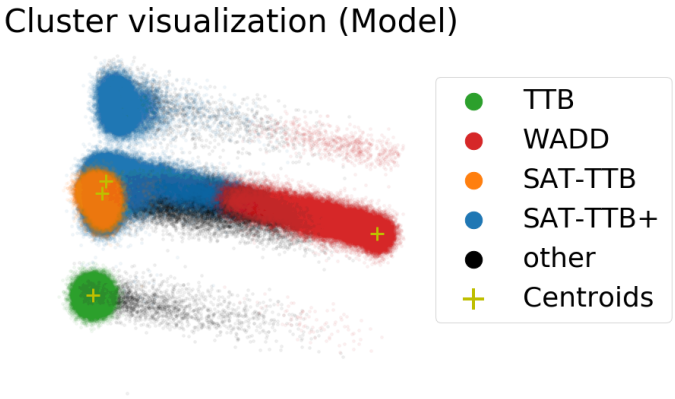

Fig. S4. Click sequence vectors from model trials projected onto the 2D LDA transform, with labels corresponding to the strategy definition.

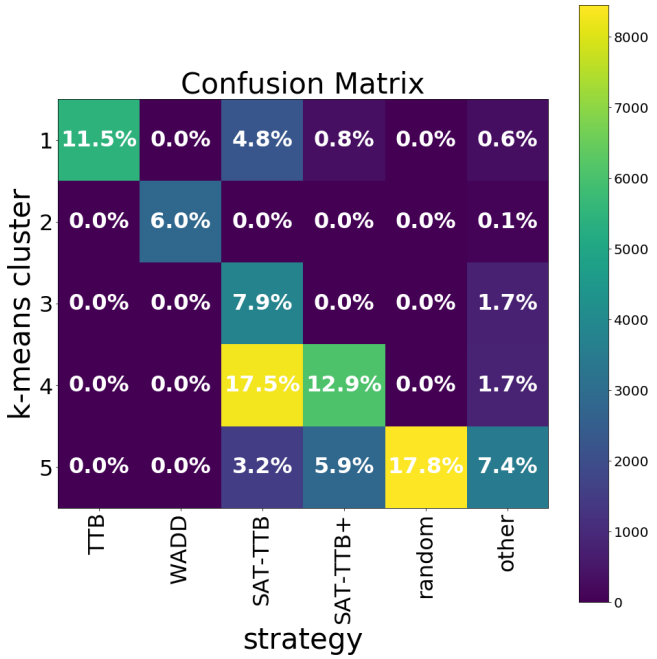

Fig. S5. Confusion matrix showing agreement between k-means cluster labels and strategy definitions, for participant trials. Annotations show the percentage of total trials accounted for by each strategy pair. Cohen's $\kappa = 0.543, 95\%$ CI $[0.537, 0.548]$

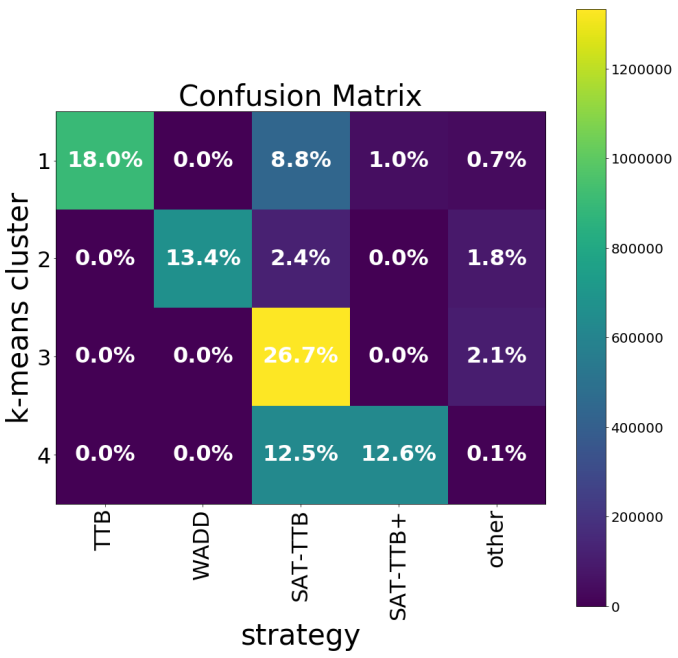

Fig. S6. Confusion matrix showing agreement between k-means cluster labels and strategy definitions, for resource-rational model trials. Annotations show the percentage of total trials accounted for by each strategy pair. Cohen's $\kappa = 0.649, 95\%$ CI $[0.648, 0.649]$

Comparison of strategies across environments

We inspected how participants adapted their strategy use frequency to the structure of the environment. Figure 4 in the main text 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.

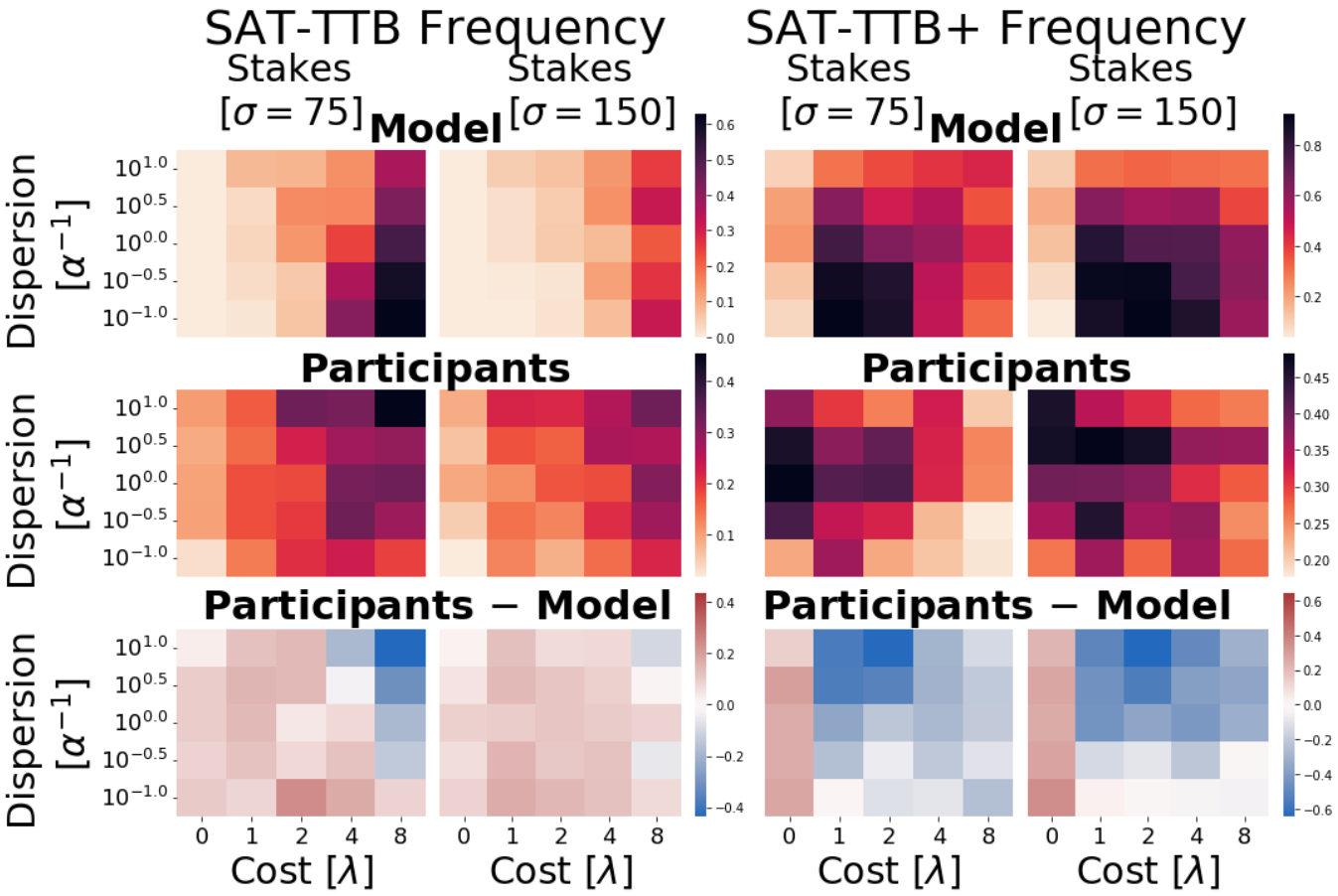

Fig. S7. Frequency of SAT-TTB+ (left panels) and SAT-TTB (right panels) across all fifty experimental conditions, for the model (top panels), participants (middle panels), and a comparison between the model and participants (bottom panels). The decision environment in each condition is defined by three parameters: σ (variance in potential reward received), α^{-1} (homogeneity of the outcome distribution), and $lambda$ (number of points deducted for each piece of information gathered). The results here accompany the results shown in Figure 4 in the main text. SAT-TTB+ and SAT-TTB are two heuristics discovered using our resource-rational method.

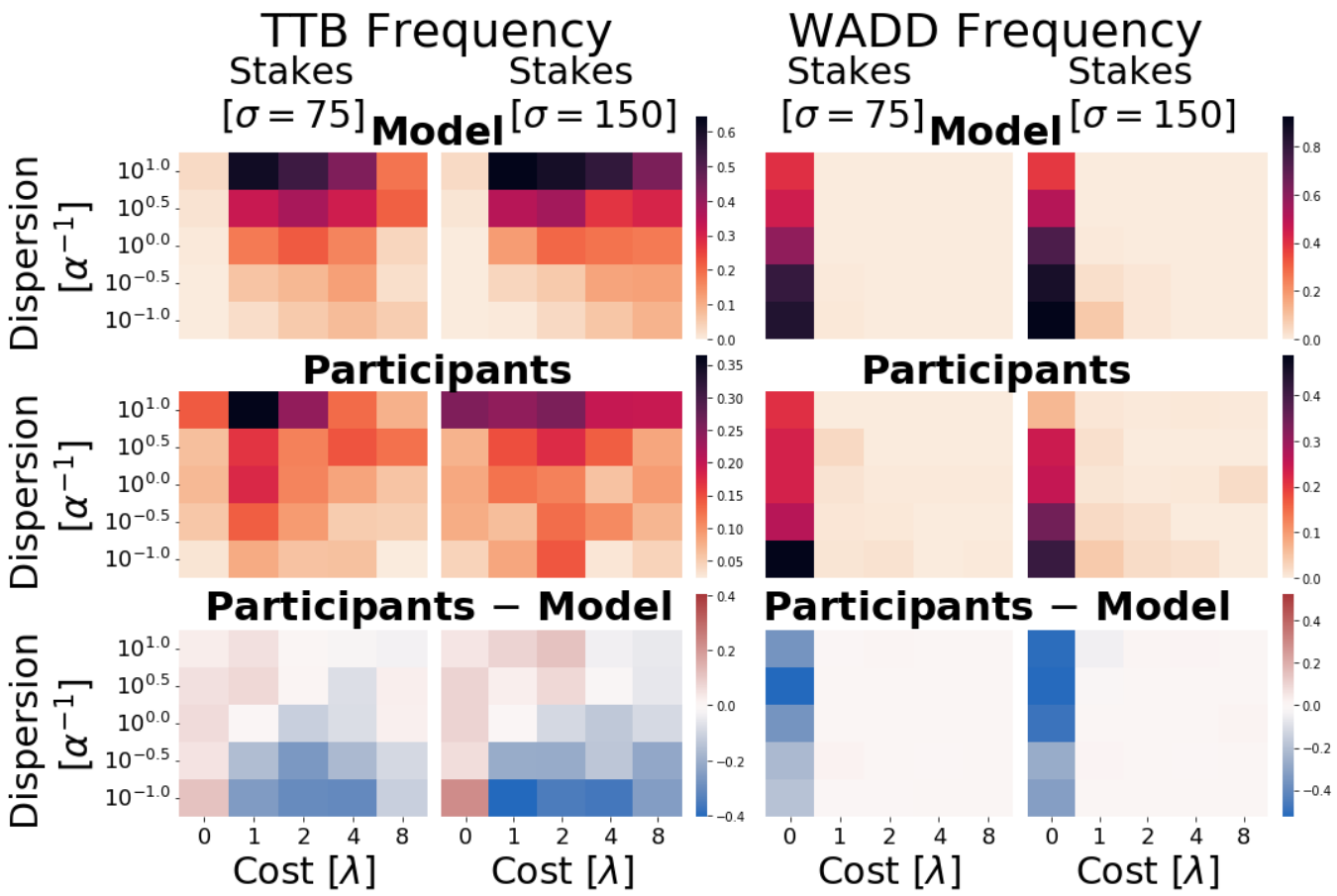

Fig. S8. TTB (left panels) and WADD (right panels) strategy use frequencies across all fifty conditions in the experiment, for the model (top panels), participants (middle panels), and a comparison between the model and participants (bottom panels). TTB and WADD are two known heuristics that our resource-rational 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 $lambda$ (number of points deducted for each piece of information gathered). This figure corresponds to Figure 4 in the main text, which shows frequencies for each parameter, collapsed across all others.

Table S1. Statistical results accompanying Fig. 4.

Strategy frequency	Independent variable	significant post-hoc comparisons	effect sizes (Cohen's d)
SAT-TTB	stakes	n/a	0.11
SAT-TTB+	stakes	n/a	0.087
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	$-0.019, -0.082, -0.076, -0.13$
TTB	cost	all pairs	$0.21, -0.047, -0.13, -0.063$
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* in the main text, and shown in Figure 4. When applicable, post-hoc pairwise comparisons were conducted between all 10 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.

Rational strategy selection explains variability in choice behavior

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 in the main text displays the main effect of each of the three parameters of the decision environment on each of these variables. Figure S9 displays these two variables in all fifty environmental conditions. Figures S10 & S11 show the alternative-variance and attribute-variance. In all cases, participants show a correspondence to the theoretical predictions of the model as to how these behavioral markers should adapt to the environment. See the *Rational strategy selection explains variability in choice behavior* subsection in the Results section of the main text for details on how these measurements were defined.

Table S2 summarizes statistical analyses accompanying those presented in the main text, corresponding to Figures 5, S4, and S6. 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 adjacent 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.

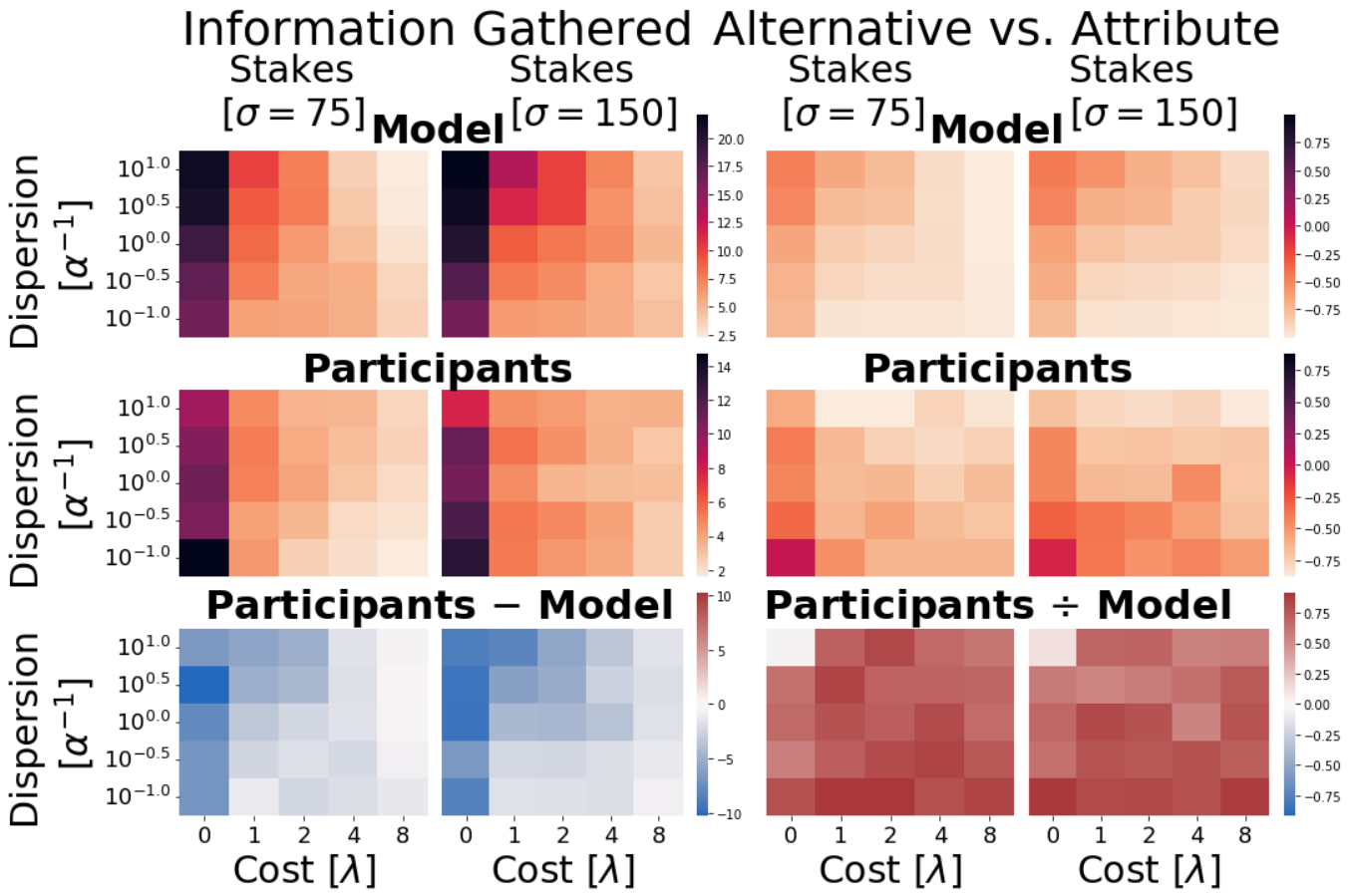

Fig. S9. Information-gathering (measured with clicks; left panels) and attribute- versus alternative-based processing (right panels) shown across all fifty conditions of the experiment, for the model (top row), human participants (middle row), and a comparison between the model and participants (bottom 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 in the main text. 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 *Rational strategy selection explains variability in choice behavior* subsection in the Results section of the main text for details on how alternative- versus attribute-based processing was measured.

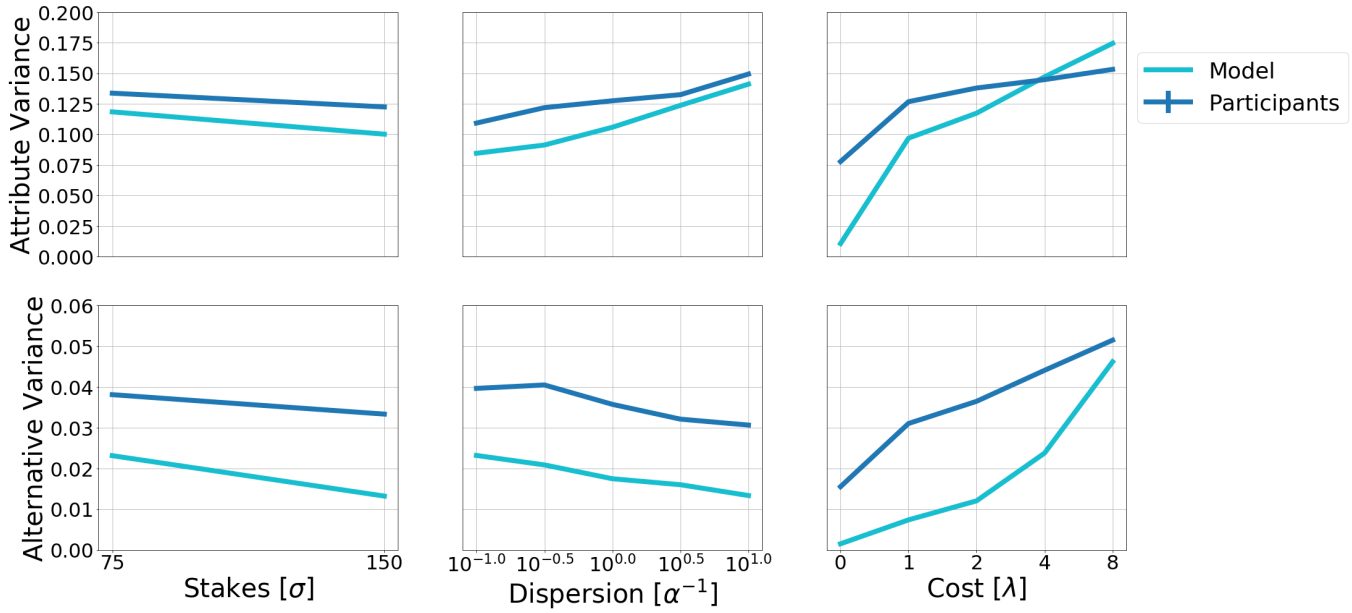

Fig. S10. Behavioral correspondence between participants and the resource-rational model. Attribute variance (top panels), and alternative variance (bottom panels) for the resource-rational model and human participants vary across the three parameters of the experiment: σ (reward stakes), α^{-1} (dispersion of outcome probabilities), and λ (cost per click). Error-bars show the SEM across participants.

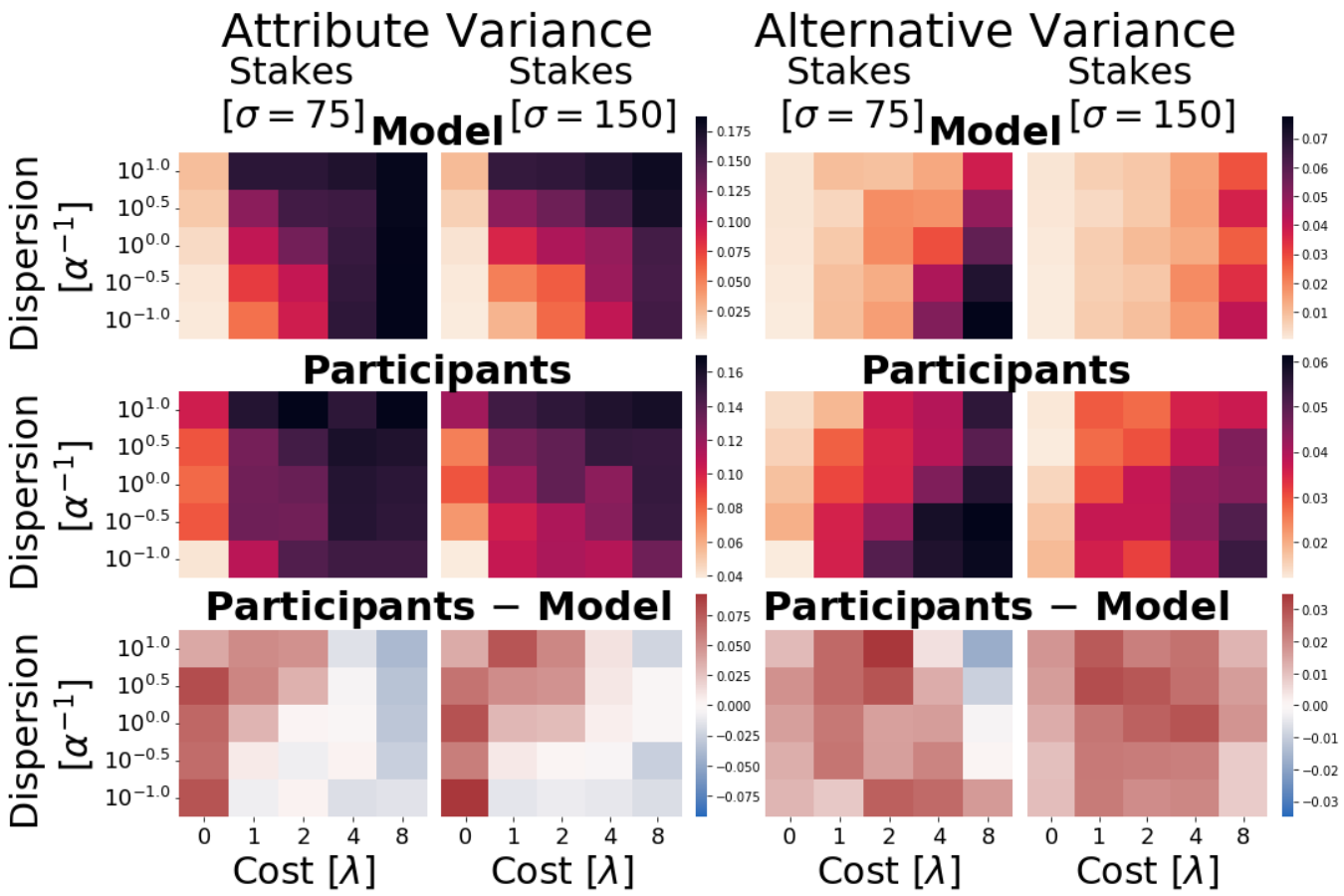

Fig. S11. Alternative and attribute variance for all fifty conditions in the experiment (all combinations of σ, α^{-1}, and $lambda$), for the model (top panels), participants (middle panels), and difference between the two (bottom panels). The results here accompany the behavioral results shown in Figure S10. Within each parameter value in Figure S10, 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 S2. Statistical results accompanying Figures 5, S4, & S6.

Dependent variable	Independent variable	main effect	significant post-hoc comparisons	effect sizes (Cohe
Information gathering	stakes	$t(2366) = 2.61$, $p = 0.009$	n/a	0.11
Information gathering	dispersion	$F(4, 2363) = -1.22$, $p = 0.30$	n/a	, -0.064, -0.0012, 0.0
Information gathering	cost	$F(4, 2363) = 293.8$, $p < 0.001$	all pairs except 2&4, 4&8	1.00, 0.32, 0.25,
Processing pattern	stakes	$t(2366) = 2.28$, $p = 0.022$	n/a	0.099
Processing pattern	dispersion	$F(4, 2363) = -28.0$, $p < 0.001$	all pairs except $10^{-0.5}$&$10^{-0.0}$, $10^{0.5}$&$10^{1.0}$	-0.16, -0.20, -0.092
Processing pattern	cost	$F(4, 2363) = 31.4$, p<0.001	0&1, 0&2, 0&4, 0&8	-0.52, -0.048, 0.012
Attribute variance	stakes	$t(2196) = -3.89$, p<0.001	n/a	-0.17
Attribute variance	dispersion	$F(4, 2193) = 24.74$, p<0.001	all pairs except $10^{-0.5}$&$10^{0.0}$, $10^{0.0}$&$10^{0.5}$	0.18, 0.010, 0.11,
Attribute variance	cost	$F(4, 2193) = 121$, p<0.001	all pairs except 2&4, 4&8	0.78, 0.20, 0.095,
Alternative variance	stakes	$t(2196) = -2.92$, p=0.0034	n/a	-0.12
Alternative variance	dispersion	$F(4, 2193) = -8.43$, 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,
Alternative variance	cost	$F(4, 2193) = 115.0$, p<0.001	all pairs	0.70, 0.24, 0.19,
Relative performance	stakes	$t(2366) = 1.69$, p=0.092, one-tailed	n/a	0.069
Relative performance	dispersion	$F(4, 2363) = -6.56$, p<0.001	$10^{-1.0}$&$10^{1.0}$, $10^{-1.0}$&$10^{0.5}$, $10^{-0.5}$&$10^{0.5}$, $10^{0.0}$&$10^{1.0}$, $10^{0.0}$&$10^{0.5}$	-0.075, 0.051, -0.28
Relative performance	cost	$F(4, 2363) = 21.0$, p<0.001	all pairs except 1&2, 2&4, 4&8	0.10, 0.11, 0.033,

Summary of statistical results corresponding to the analyses shown in Figures. 5 & S1. 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 adjacent 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.

Sources of under-performance

We assessed how close human performance comes to the upper bound established by the performance of our resource-rational model. We measured people's relative performance by the fraction of the highest expected reward attainable with perfect information, and omitting the cost of information gathering. Following the predictions of our resource-rational model, participants' relative performance tended to increase with increasing stakes (standardized $\beta = 0.00023, p = 0.092$) and with increasing dispersion of the outcome distribution ($\beta = 0.0048, p < 0.001$), and decrease with increasing click costs ($\beta = -0.010, p < 0.001$; Figure S12).

To evaluate the degree to which participants' decision strategies are resource-rational, we measured the performance of participants versus resource-rational decision-making. This analysis revealed that, on average, the relative performance of our participants' decision strategies was 60.9% of the relative performance of resource-rational decision-making (and 71.2% when excluding participants who gamble randomly on more than half of all trials). There are at least four possible reasons why people might be suboptimal: implicit costs of information gathering, suboptimal use of the gathered information, suboptimal strategy selection, and suboptimal strategy execution. We now assess the degree to which each of these contributes to people's under-performance in turn.

To assess the degree to which insufficient information gathering led to participants' suboptimal performance, we ran 1,000 simulations of our method on each of the same exact trials presented to human participants, and measured relative performance (defined as the fraction of the highest possible reward attainable with perfect information, and omitting the cost of information gathering). To control for the overall amount of information gathered between our method and participants, we fit an implicit cost of information gathering to the model using a grid search and found that an implicit cost of 2.1 points per click led to the

same amount of information gathering on average as participants. We then measured the relative performance of the model with an implicit cost of clicking of 2.1. This procedure was repeated when omitting participants who gambled randomly on more than half of trials, resulting in an implicit cost of clicking of 1.3 points per click.

We then compared human performance to the performance of the resource-rational agent with an additional implicit cost for each cell revealed, set such that the agent gathered on average the same amount of information as people. Such an implicit cost might capture, for example, the physical effort and time required for humans to click (see Materials and Methods, and Discussion sections). When the implicit cost was 2.1 points per click our model gathered the same amount of information as participants on average, yet this resulted in only a 9.0% decrease in the model's relative performance, accounting for 23.0% of the overall difference in the relative performance between the perfectly resource-rational model and participants (see the dashed teal lines in Figure S12; see Figure S14 for a comparison of the relative performance of participants and the model with and without an implicit cost, across all fifty experimental conditions).

To assess the degree to which suboptimal strategy selection and suboptimal strategy execution led to under-performance, we calculated a confusion matrix of strategy selection for participants and the model, using the same simulations. Since the method occasionally selects different strategies for the same trial (depending on the location of the first click), for a given trial, the strategy selected on each simulation was counted as $1/1{,}000^{\text{th}}$ of a trial, to account for the full distribution of strategy selections. This procedure was repeated for the model without (Figure S16) and with (Figure S18) an implicit cost. For each cell in the confusion matrix, we calculated the percentage of the point differential between model and participants accounted for by the corresponding trials (where the point differential was for suboptimal strategy selection and suboptimal strategy execution only, that is, not including the point differential accounted for by implicit costs or suboptimal strategy execution). These percentages are displayed as annotations in Figures S16 and S18. In Figures S17 and S19 we displayed the same confusion matrices shown in Figures S16 and S18, respectively, but only including the four strategies deployed by the model and participants, and normalizing the colormap for each row for visualization purposes, while displaying annotations for the absolute point differential accounted for by each cell. To quantify the agreement in each confusion matrix, we calculated Cohen's kappa (**?**).

To measure the extent to which people are suboptimal because they make imperfect use of the collected information, we computed the subjective expected values of all alternatives according to the information revealed by the participant. We found that participants chose the gamble with the highest subjective expected value 90.2% of the time, accounting for only a 4.2% reduction in participants' relative performance (an average loss of 2.76 points per trial), or 6.8% of the overall difference between model and participant relative performance.

To assess whether participants' under-performance was due to suboptimal strategy selection and suboptimal strategy execution, we constructed a confusion matrix of trial-wise strategy selection for the model and participants, separately for the model without (Figures S16 and S17) and with (Figures S18 and S19) the implicit cost of clicking. We then summarized overall agreement in strategy selection with Cohen's κ. Overall, the agreement between the strategies selected by our model and participants on a trial-by-trial basis was significantly above chance ($p < 0.001$) but indicated rather low agreement for the top four strategies ($\kappa = 0.16$ without implicit cost, $\kappa = 0.11$ with implicit cost). The annotations in Figure S16 show that, when controlling for the implicit cost of clicking (and assuming perfect participant strategy execution to isolate the effect of strategy selection), trials in which the model selects SAT-TTB+ and participants gamble randomly account for 34.7% of the total contribution of suboptimal strategy selection to reduced relative performance, and other trials when the model selects SAT-TTB+ or when participants gamble randomly account for another 58.1%.

To calculate the degree to which suboptimal strategy execution led to under-performance, we calculated how many points were lost by not selecting the gamble with the highest expected value (conditioned on the information gathered).

Suboptimal strategy selection—when participants choose a different strategy than the model on a given trial—accounts for 63.0% of under-performance, while suboptimal strategy execution—when participants use the optimal strategy on a given trial but do not gather information from the best locations—accounts for the remaining 7.2% of the performance gap between people and the model.

Table S3 summarizes the contributions of each of these four sources of under-performance to absolute and proportional loss in relative performance. It also presents the contribution of these sources of under-performance when excluding participants who gambled randomly on more than half of all trials (394 participants, or 16.6% of all participants). With this exclusion criterion (corresponding to Figure S13), the absolute contribution of suboptimal information gathering to performance is cut roughly in half, and the implicit cost of clicking is reduced from 2.1 to 1.3 points per click. The absolute contribution of suboptimal strategy selection is reduced by a quarter, and still accounts for the majority of overall 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.

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; $\beta = -0.62, t(38737) = -37.6, p < 0.001$).

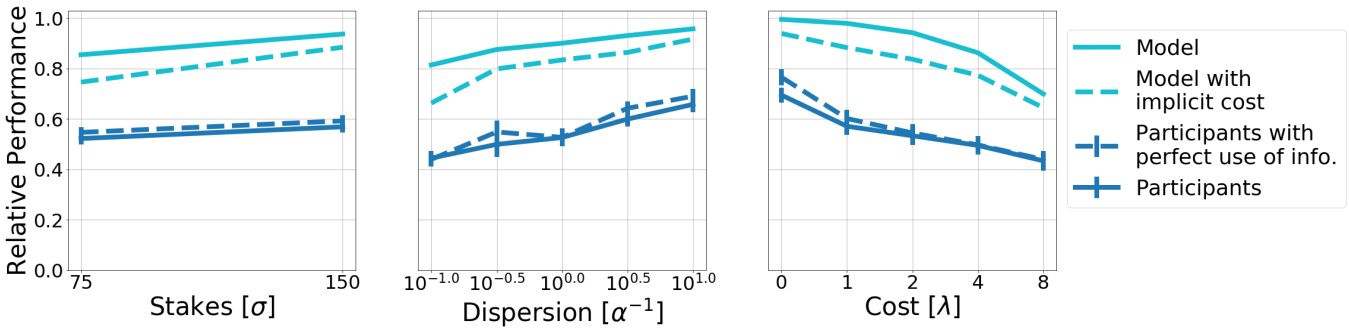

Fig. S12. Participants show a qualitative correspondence in relative performance across conditions. Performance was measured as the relative reward earned on each trial (the fraction of the highest possible reward with perfect information, omitting click costs). The dashed lines show the relative performance of the model when it is charged an additional 2.1 points per click. This represents an implicit cost of clicking not captured by our model, such that the model gathers the same amount of information on average as participants. This implicit cost reduced the relative performance of the model by only 9.0%. See the *Rational strategy selection explains variability in choice behavior* subsection in the Results section of the main text for more information. Error-bars show the SEM across participants.

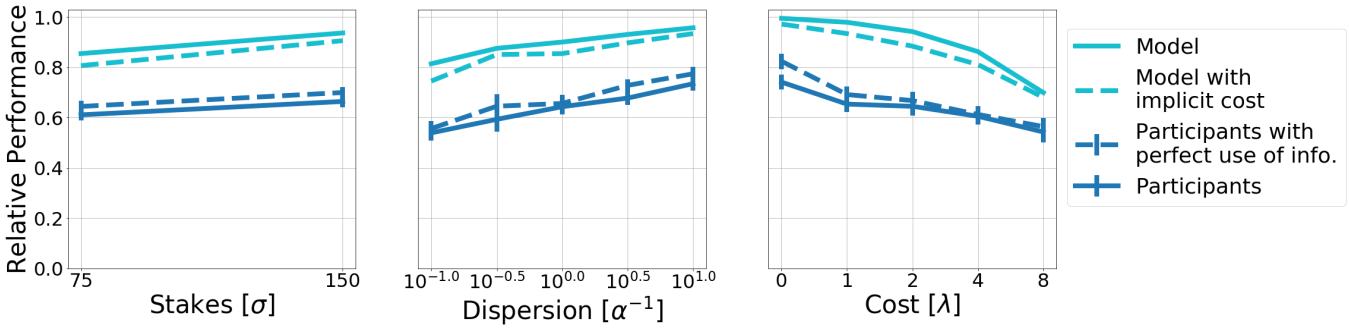

Fig. S13. Same figure as above, but excluding participants who gambled randomly on over half of all trials ($n = 394$ of $2,368$ participants total). This led to a reduced implicit cost of clicking of 1.3 points per click. See Table S3 for a comparison of sources of under-performance when either excluding these participants or not.

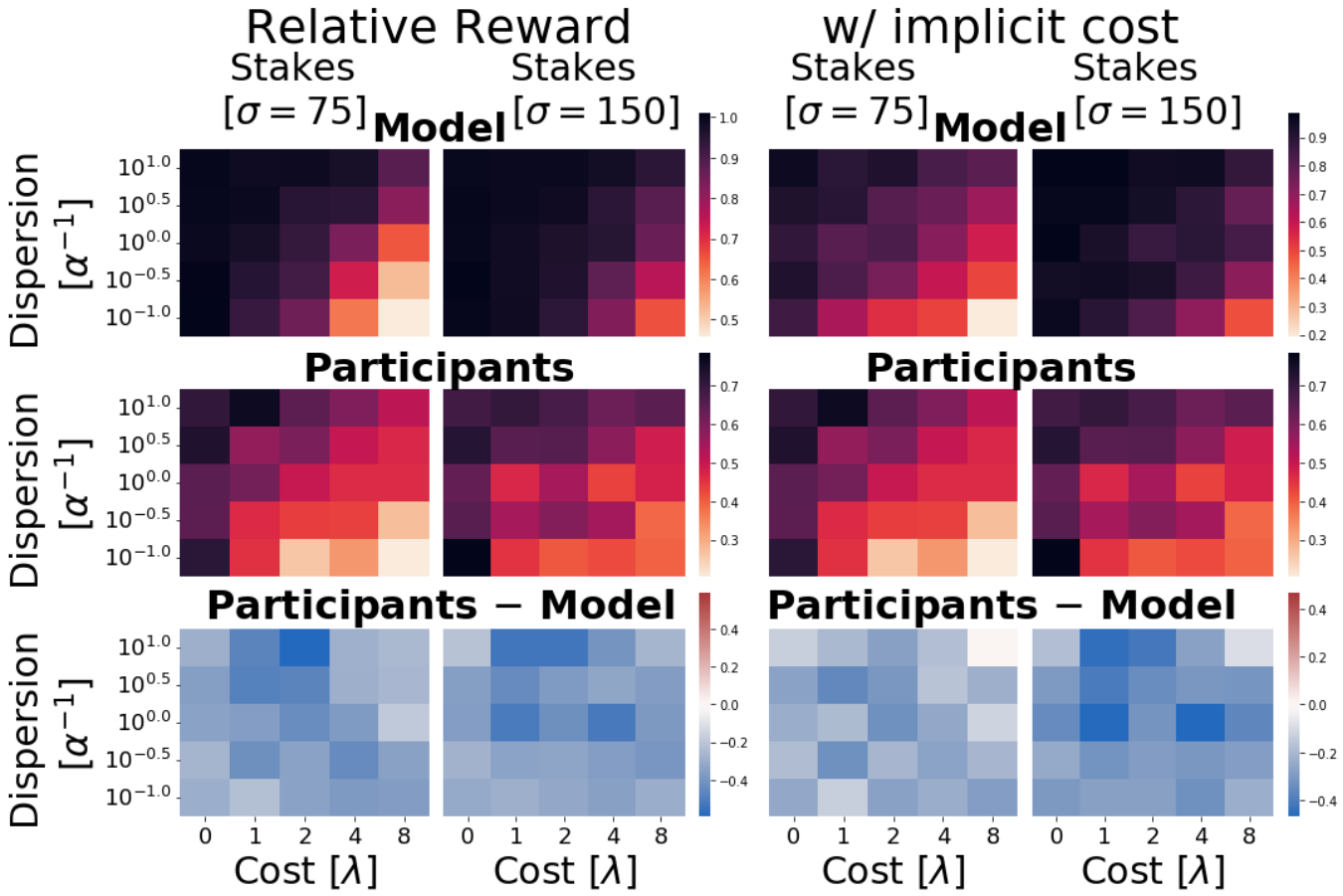

Fig. S14. Relative performance (right panels) and relative performance for the model with an implicit cost of clicking (left panels) shown across all fifty conditions of the experiment, for the model (top row), human participants (middle row), and the difference between the model and participants (bottom 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 S12. Within each parameter value in Figure S12, results are averaged across all values of other parameters, whereas in this figure the full results for each of the fifty conditions is shown.

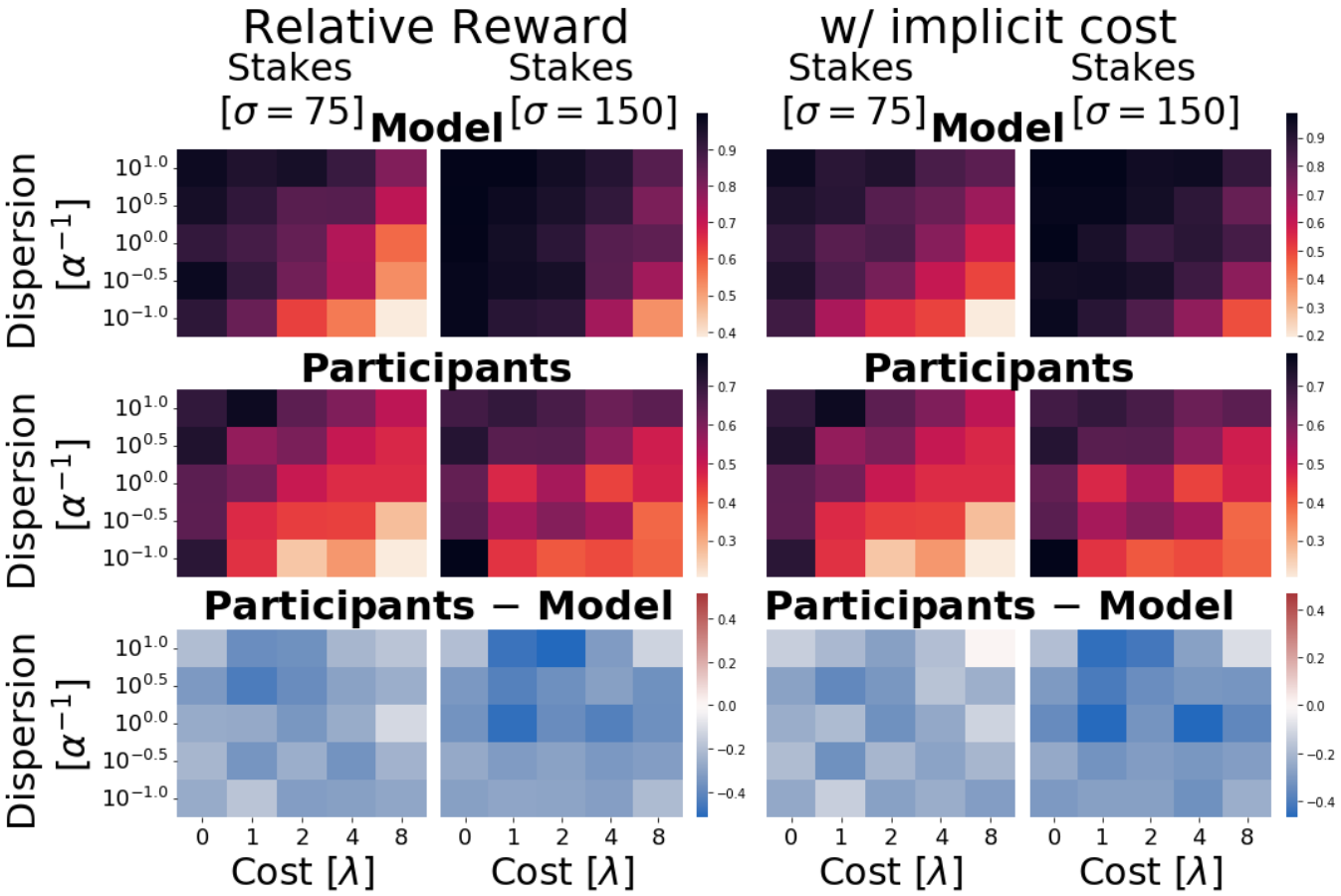

Fig. S15. Same as Figure S14, but excluding participants who gambled randomly on more than half of all trials ($n = 394$ of $2{,}368$ participants total).

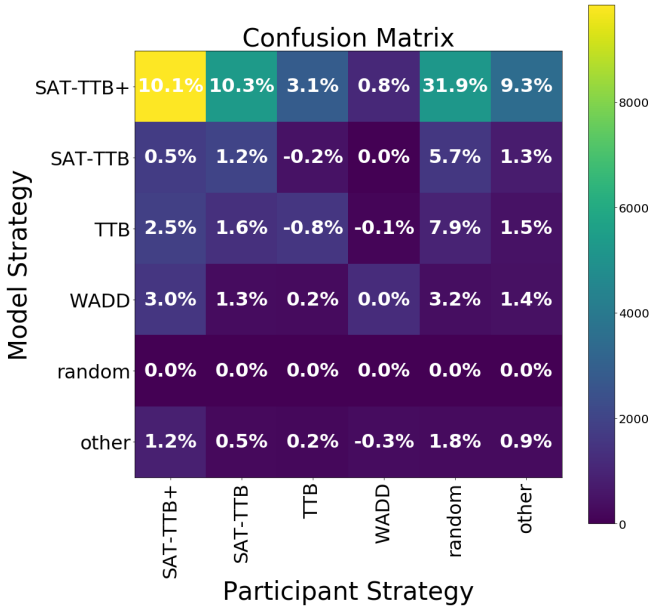

Fig. S16. Confusion matrix showing the number of trials in agreement in strategy selection between the model and human participants. Annotations show the percentage of the total difference in relative performance between the model and participants accounted for by each strategy pair (assuming perfect participant strategy execution). Cohen's $\kappa = 0.087, 95\%$ CI $[0.082, 0.091]$

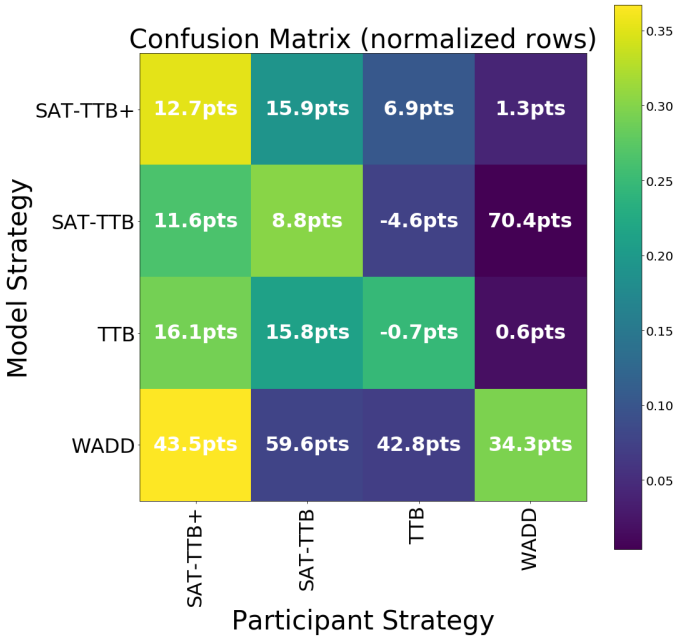

Fig. S17. Confusion matrix showing the portion of trials in agreement in strategy selection between the model and human participants, normalized by model strategy selection frequency (rows), showing the top four strategies. Annotations show the average point difference between the model and participants for each strategy pair. Cohen's $\kappa = 0.16, 95\%$ CI $[0.15, 0.17]$

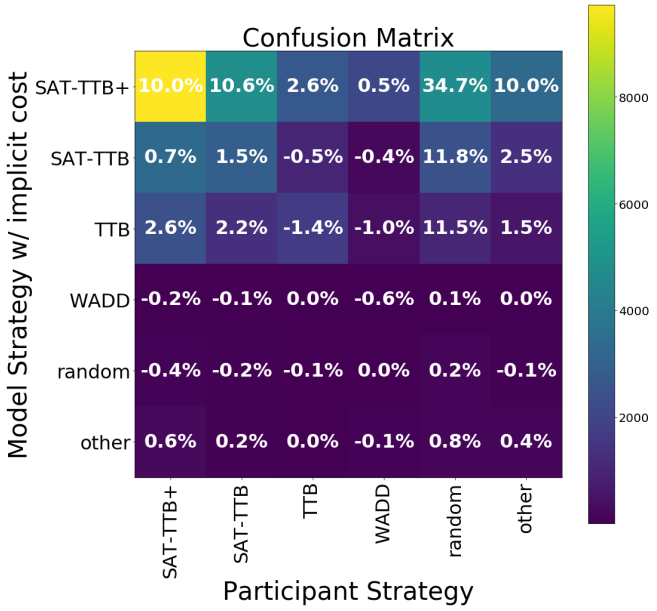

Fig. S18. Confusion matrix showing the number of trials in agreement in strategy selection between the model and human participants, where the model used the same amount of information gathering on average as participants (with an implicit cost of clicking of 2.1 points per click). Annotations show the percentage of the total difference in relative performance between the model and participants accounted for by each strategy pair (assuming perfect participant strategy execution). Controlling the amount of information gathered for the model and participants, plus assuming perfect strategy execution for participants, allows us to isolate the effect of strategy selection on relative performance. Cohen's $\kappa = 0.067, 95\%$ CI $[0.062, 0.072]$

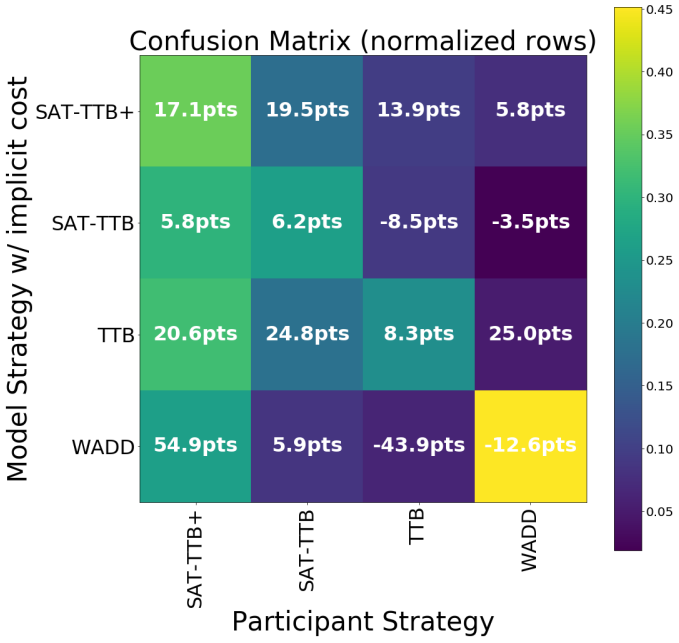

Fig. S19. Confusion matrix showing the portion of trials in agreement in strategy selection between the model and human participants, normalized by model strategy selection frequency (rows), showing the top four strategies. The model used the same amount of information gathering on average as participants. Annotations show the average difference between model and participant net points per trial, for each strategy pair. Cohen's $\kappa = 0.11, 95\%$ CI $[0.10, 0.12]$

Table S3. Contributions of three sources of participant under-performance

Source of under-performance	Participant exclusion	Absolute loss of relative performance	Proportion of model-participant differe in relative performance
Implicit costs	No	0.080	23.0%
Suboptimal information use	No	0.024	6.8%
Suboptimal strategy selection	No	0.22	63.0%
Suboptimal strategy execution	No	0.025	7.2%
Implicit costs	Yes	0.039	15.2%
Suboptimal information use	Yes	0.034	13.4%
Suboptimal strategy selection	Yes	0.16	62.2%
Suboptimal strategy execution	Yes	0.024	9.2%

Suboptimal strategy selection accounts for the majority of participant under-performance (see main text for details of these analyses). Performance is defined using relative reward, as in Figures S12 and S13. Implicit costs of clicking account for a reduction of just 0.080 units of relative reward (the average difference between the dashed and solid teal lines in Figure S12), or 23.0% of the total difference in relative performance between the model and participant (i.e. 23.0% of the average difference in the solid teal and solid blue lines in Figure S12, 60.9%). Suboptimal use of information accounts for a reduction of a mere 0.024 units of reward in performance (the average difference between the dashed and solid blue lines in Figure S12), or 6.8% of the 60.9% overall difference in performance between the model and participants. Suboptimal strategy selection accounts for a reduction of 0.22 units of reward (the average difference between the dashed teal and dashed blue lines in Figure S12), which is 63.0% of the overall difference in reward between the model and participants. Suboptimal strategy execution accounts for the remaining 7.2% reduction in performance. The annotations in Figure S18 shows the contribution of every pair of model-participant strategy types, averaged across trials, to the difference in performance between the model and participants, controlling for the amount of information gathered. That is, the annotated percentages in Figure S18 (which add up 100%) are percentages of 70.2%, the contribution of suboptimal strategy selection and execution to overall under-performance. The annotated percentages in Figure S16 (which also add up to 100%) do not control for the amount of information gathered between the model and participants, and therefore account for suboptimal strategy selection and execution, and implicit costs (93.2% of the overall difference in relative performance between the model and participants). When excluding participants who gamble randomly on more than half of all trials ($n = 394$ of 2,368 participants total), then performance at the group level goes from 60.9% to 71.2% of the resource-rational model, and the contribution to under-performance of suboptimal information gathering goes down (with the implicit cost of clicking reducing from 2.1 to 1.3 points per click).