# When to use which heuristic: A rational solution to the strategy selection problem

**Falk Lieder (falk.lieder@berkeley.edu)**
Helen Wills Neuroscience Institute
University of California at Berkeley, CA, USA

**Thomas L. Griffiths (tom_griffiths@berkeley.edu)**
Department of Psychology
University of California at Berkeley, CA, USA

**Abstract**

The human mind appears to be equipped with a toolbox full of cognitive strategies, but how do people decide when to use which strategy? We leverage rational metareasoning to derive a rational solution to this problem and apply it to decision making under uncertainty. The resulting theory reconciles the two poles of the debate about human rationality by proposing that people gradually learn to make rational use of fallible heuristics. We evaluate this theory against empirical data and existing accounts of strategy selection (i.e. SSL and RELACS). Our results suggest that while SSL and RELACS can explain people's ability to adapt to homogeneous environments in which all decision problems are of the same type, rational metareasoning can additionally explain people's ability to adapt to heterogeneous environments and flexibly switch strategies from one decision to the next.

**Keywords:** Strategy Selection; Decision Making; Heuristics; Bounded Rationality; Cognitive Control; Learning

## Introduction

Many of our decisions and judgments systematically violate the laws of logic and probability theory (Tversky & Kahneman, 1974). These violations are known as cognitive biases. Cognitive biases have been interpreted as evidence that people do not reason by the rules of logic and probability theory but by simple yet fallible heuristics. Whether or not these findings prove that humans are irrational has been debated for decades (Stanovich, 2009). While some view heuristics and biases as a sign of human irrationality, recent work suggests that some heuristics can be understood as rational strategies once computational costs are taken into account (Lieder, Griffiths, & Goodman, 2013; Griffiths, Lieder, & Goodman, 2015; Lieder, Hsu, & Griffiths, 2014). Furthermore, Gigerenzer and colleagues argue that having a toolbox of simple heuristics that are well adapted to the structure of our environment makes us smart (Gigerenzer & Todd, 1999).

Yet, being a skilled carpenter requires more than a toolbox: you also have to know when to use which tool. Todd and Gigerenzer (2012) postulate that we choose heuristics that are well-adapted to our current situation (i.e. *ecologically rational*), but they do not explain how we are able to do so. Empirical evidence suggests that people do indeed choose heuristics adaptively (Payne, Bettman, & Johnson, 1988; Bröder, 2003; Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011). Despite some progress, the computational principles of strategy selection remain unclear (Marewski & Link, 2014). Previous theories of strategy selection, namely SSL (Rieskamp & Otto, 2006), RELACS (Erev & Barron, 2005), and SCADS (Shrager & Siegler, 1998) predict the formation of rigid mental habits that always pursue the same strategy, whereas people are more flexible (Lieder, Plunkett, et al., 2014; Payne et al., 1988).

This paper proposes a rational solution to the strategy selection problem: in analogy to model-based reinforcement learning (Dolan & Dayan, 2013) our theory posits that people learn a mental model that enables them to predict each heuristic's accuracy and execution time from features of the problem to be solved and choose the heuristic with the best predicted speed-accuracy tradeoff.

In the remainder of this paper we first review previous strategy selection theories and introduce our new theory of strategy-selection. We then test our theory against those previous accounts by fitting published data on multi-attribute decision making, conducting a novel experiment, and demonstrating that our theory can account for people's adaptive flexibility in risky choice. We close with a discussion of the implications of our results for the debate about human rationality and directions for future research.

## Models of Strategy Selection

According to previous theories of strategy selection we learn to choose the strategy that works best *on average* across all problems in an environment (Rieskamp & Otto, 2006; Erev & Barron, 2005) or category (Shrager & Siegler, 1998). This approach ignores that every problem has distinct characteristics that determine the strategies' effectiveness. After reviewing these *context-free* theories, we propose a model that chooses strategies based on the features of individual problems.

### Context-free strategy selection learning

According to the SSL model (Rieskamp & Otto, 2006) the probability that strategy $s$ will be chosen ($P(S_t = s)$) in trial $t$ is proportional to its reward expectancy $q_i$:

$$P(S = s) \propto q_t(s), \tag{1}$$

where $q_t(k)$ is the sum of the rewards obtained when strategy $k$ was chosen prior to trial $t$ plus the initial reward expectancy

$$q_0(k) = r_{\max} \cdot w \cdot \beta_k, \tag{2}$$

where $r_{\max}$ is the highest possible reward, $w$ is the strength of the initial reward expectancy, and $\beta_1, \cdots, \beta_N \in [0,1]$ are the agent's initial relative reward expectancies for strategies $1, \cdots, N$ and sum to one.

The RELACS model (Erev & Barron, 2005) chooses strategies according to their recency-weighted average payoffs

$$W_{t+1}(k) = \begin{cases} \alpha \cdot r_t + (1-\alpha) \cdot W_t(k) & \text{if } S_t = k \\ W_{t+1}(k) = W_t(k) & \text{else} \end{cases} \tag{3}$$

$$P(S_t = k) \propto e^{\lambda \cdot \frac{W_t(k)}{V_t}} \tag{4}$$

where the parameters $\alpha$ and $\lambda$ determine the agent's learning rate and decision noise respectively, and $V_t$ is the agent's current estimate of the payoff variability.

The SCADS model (Shrager & Siegler, 1998) presupposes that each problem has been identified as an instance of one or more problem types and assumes associative learning mechanisms similar to those of SSL.

### Feature-based strategy selection learning

In this section, we present a theory according to which people learn a mental model predicting the effectiveness of cognitive strategies from features of the problem to be solved.

Strategy selection is a metacognitive decision with uncertain consequences. We therefore leveraged *rational metareasoning* – a decision-theoretic framework for choosing computations (Russell & Wefald, 1991) – to develop a rational model of strategy selection that is theoretically sound, computationally efficient, and competitive with state-of-the-art algorithm selection methods (Lieder, Plunkett, et al., 2014).

Rational metareasoning chooses the strategy $s^\star$ with the highest value of computation (VOC) for the problem specified by input $\mathbf{i}$:

$$s^\star = \arg\max_{s \in \mathcal{S}} \mathrm{VOC}(s, \mathbf{i}), \tag{5}$$

where $\mathcal{S}$ is the set of the agent's cognitive strategies. The VOC of executing a cognitive strategy $s$ is the expected net increase in utility over acting without deliberation. If the strategy chooses an action and the utility of the available actions remains approximately constant while the agent deliberates, then the VOC can be approximated by the expected reward of the resulting action minus the opportunity cost of the strategy's execution time $T$:

$$\mathrm{VOC}(s; \mathbf{i}) \approx \mathbb{E}\left[R | s, \mathbf{i}\right] - \mathbb{E}\left[\mathrm{TC}(T) | s, \mathbf{i}\right], \tag{6}$$

where $R$ is the increase in reward and $\mathrm{TC}(T)$ is the opportunity cost of running the algorithm for $T$ units of time. The reward $R$ can be binary (correct vs. incorrect output) or numeric (e.g., the payoff). Equations 5-6 reveal that near-optimal strategy selection can be achieved by learning to predict the strategies' expected rewards and execution times from features $\mathbf{f}(\mathbf{i})$ of the input $\mathbf{i}$ that specifies the problem to be solved. These predictions can be learned by Bayesian linear or logistic regression as described in Lieder, Plunkett, et al. (2014).

Equation 5 is optimal when the VOC is known, but when the VOC is unknown the value of exploration should not be ignored. To remedy this problem, we employ Thompson sampling (Thompson, 1933) – a near optimal solution to the exploration-exploitation dilemma (May, Korda, Lee, & Leslie, 2012). Concretely, each strategy $s$ is chosen ($S = s$) according to the probability that its VOC is maximal:

$$P(S = s) \propto P\left(s = \arg\max_{s} \mathrm{VOC}(s; \mathbf{i})\right). \tag{7}$$

This is implemented by drawing one sample from each strategy's VOC model and picking the strategy with the highest sampled value (ties are broken at random). This proposal is line with behavioral (Vul, Goodman, Griffiths, & Tenenbaum, 2014) and neural evidence (Fiser, Berkes, Orbán, & Lengyel, 2010) for sampling as a cognitive mechanism.

## Learning when to use fast-and-frugal heuristics

Fast-and-frugal heuristics perform very few computations and use only a small subset of the available information (Gigerenzer & Gaissmaier, 2011). For instance, the Take-the-Best (TTB) heuristic for multi-attribute decision making chooses the option with the highest value on the most predictive attribute that distinguishes the options and ignores all other attributes. This strategy works in so-called *non-compensatory* environments in which the attributes' predictive validities fall off so rapidly that the recommendation of the most predictive attribute cannot be overturned by rationally incorporating other attributes. Yet it can fail miserably in compensatory environments in which no single attribute reliably identifies the best choice by itself.

Bröder (2003) found that people use Take-the-Best more frequently in non-compensatory environments than in compensatory environments. Rieskamp and Otto (2006) conducted an experiment suggesting that this adaptation might result from reinforcement learning: Two groups of participants made 168 multi-attribute decisions with feedback in a compensatory versus a non-compensatory environment. Over time, the choices of participants in the non-compensatory environment became more consistent with TTB, whereas the choices of participants in the compensatory environment became less consistent with TTB and more consistent with the weighted-additive strategy (WADD) that computes the weighted average of all attributes.

These findings raise the question how people learn when to use TTB. This problem could be solved either by learning how well TTB works *on average*, as postulated by SSL and RELACS, or by learning to predict the performance of TTB and alternative strategies for individual problems as suggested by rational metareasoning.

As a first test of our model we demonstrate that it can explain the findings of Experiment 1 from Rieskamp and Otto (2006). This experiment was structured into seven blocks comprising 24 trials each. In each trial participants chose between two investment options based on five binary attributes whose predictive validities were constant and explicitly stated. To apply our general rational metareasoning theory to multi-attribute decision making, we manually selected a small set of simple features that are highly informative about the relative performance of TTB versus WADD: The first feature predicts the performance of TTB by the validity of the most reliable discriminative attribute ($f_1$), the second and the third feature measure the potential for WADD to perform better than TTB by the gap between the validity of the most reliable attribute favoring the first option and the most reliable attribute favoring the second option ($f_2$) respectively the absolute difference between the number of attributes fa-

voring the first option and the second option respectively ($f_3$). The simulated agent's toolbox contained two strategies: Take-The-Best ($s_1 = \text{TTB}$) and the weighted-additive strategy ($s_2 = \text{WADD}$). The probability that strategy $s$ leads to the correct decision was modeled by

$$P(R=1|s) = \frac{1}{1+\exp(-\mathbf{f}\cdot\alpha_s + b_s)}. \quad (8)$$

To accommodate people's prior knowledge about the strategies' performance we parameterized its prior distribution by

$$P(\alpha_s) = \mathcal{N}\left(\mu = \mathbf{0}, \Sigma^{-1} = \tau\cdot\mathbf{I}\right) \quad (9)$$

$$P(b_s) = \mathcal{N}(\mu = b_0^{(s)}, \sigma^{-2} = \tau). \quad (10)$$

where $b_0$ and $\tau$ are free parameters and $\mathbf{0}$ and $\mathbf{I}$ are the zero vector and the identity matrix respectively.

We created compensatory and non-compensatory environments similar to those used by Rieskamp and Otto (2006): In the non-compensatory environment TTB always makes the Bayes-optimal decision, and in the compensatory environment WADD always makes the Bayes-optimal decision. In both environments TTB and WADD make the same decision on exactly half of the trials. To determine the Bayes-optimal decision and generate payoffs, we computed the probability that option A is superior to option B by Bayesian inference under the assumption that their order is uninformative and that positive and negative ratings are equally common. Payoffs were sampled from the posterior distribution given the attributes' values and validities. Furthermore, we assumed that the participants' opportunity cost $c$ corresponded to \$5 per hour at 1 computation per second. Since the initial bias in strategy selection only depends on the difference between the prior beliefs about the two strategies, we set $b_0^{(\text{TTB})}$ to zero and fit $\Delta b_0 = b_0^{(\text{WADD})} - b_0^{(\text{TTB})}$ and $\tau$ to the average frequency with which Rieskamp and Otto's participants used TTB versus WADD in each block by minimizing the mean squared error using grid search. For each grid point we simulated trial-by-trial learning and decision making and averaged the choice frequencies within each block across 1000 simulations. The resulting parameter estimates were $\Delta\hat{b}_0 = 0.14$ and $\tau = 80$.

We found that rational metareasoning can explain people's ability to adapt to compensatory as well as non-compensatory environments (see Figure 1): When the environment was non-compensatory rational metareasoning learned to use TTB, but when the environment was non-compensatory rational metareasoning learned to avoid TTB and use WADD instead. Our simulation results show that rational metareasoning captured that people gradually adapt their strategy choices to the decision environment. The fits of rational metareasoning and the fit of SSL reported by Rieskamp and Otto (2006) were about equally good (MSE: 0.0050 vs. 0.0048; see Figure 1).

## Strategy selection in mixed environments

Since both SSL and rational metareasoning can explain the results of Rieskamp and Otto (2006), a new experiment is needed to determine if strategy selection learning is context-free as postulated by SSL and RELACS or feature-based as postulated by rational metareasoning. We thus investigated under which conditions rational metareasoning predicts different strategy choices than SSL and RELACS: We evaluated the performance of context-free versus feature-based strategy selection learning in 11 environments with $p \in \{0\%, 10\%, 20\%, \cdots, 100\%\}$ compensatory problems (in which WADD makes the right and TTB makes the wrong decision) and $1-p \in \{100\%, 90\%, 80\%, \cdots, 0\%\}$ non-compensatory problems (in which TTB makes the right and WADD makes the wrong decision). Each environment comprised 168 decision problems in random order. We evaluated the performance of rational metareasoning with $b_0 = 0$ and $\tau = 1$, SSL with $\beta_1 = \beta_2 = 0.5$ and $w = 1$, and RELACS with $\alpha = 0.1$ and $\lambda = 1$. These parameters correspond to a weak bias towards using both strategies equally often, but this is not critical since any bias is eventually overwritten by experience. Our simulations revealed that the variance $p\cdot(1-p)$ of the problems' compensatoriness has qualitatively different effects on the performance of feature-based versus context-free strategy selection learning; see Figure 2. Concretely, the performance of context-free strategy selection learning drops rapidly with the variance in the environment's compensatoriness: As the ratio of compensatory to non-compensatory problems approaches $50/50$ the performance of SSL and RELACS drops to the chance level. The performance of rational metareasoning, by contrast, is much less susceptible to the variance and stays above 70%. The reason for this difference is that rational metareasoning learns to use TTB for non-compensatory problems and WADD for compensatory problems whereas SSL and RELACS learn to always use the same strategy. Rational metareasoning outperforms RELACS across all environments. In purely (non)compensatory environments SSL and rational metareasoning perform equally well, but as the environment becomes variable the performance of SSL drops below the performance of rational metareasoning. Thus context-free and feature-based strategy selection make qualitatively different predictions about people's

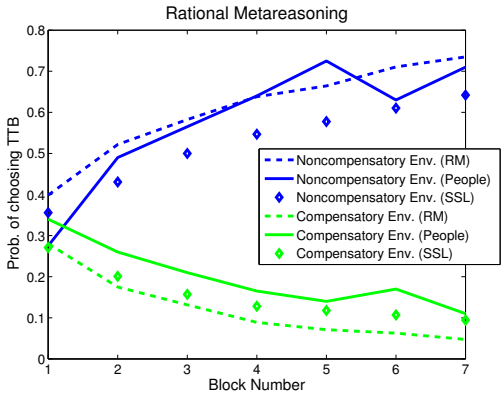


Figure 1: Rational metareasoning explains experimental findings by Rieskamp and Otto (2006).

performance in mixed environments. We can therefore determine if people use context-free or feature-based strategy selection by measuring their performance in a mixed environment with the following experiment:

## Methods

We recruited 120 participants on Amazon Mechanical Turk. Each participant was paid 50 cents for about five minutes of work. The experiment comprised 30 binary decisions. Participants played the role of a banker deciding which of two companies receives a loan based on the companies' ratings on six criteria. The criteria and their success probabilities were the same as in Rieskamp and Otto's first experiment. On each trial, the two companies' ratings on these criteria were presented in random order, and the criterias' validities were stated explicitly. After choosing company A or company B participants received stochastic binary feedback generated according to the cue validities. The relative frequency of positive feedback was 75.96% following the correct response and 25.04% following the incorrect response. The decision problems were chosen such that TTB and WADD make opposite decisions on every trial. In half of the trials, the decision of TTB was correct and in half of the trials the decision of WADD was correct. Thus, always using TTB, always using WADD, choosing one of the two strategies at random, or context-free strategy selection would result in an accuracy of 50%; see Figure 2.

## Results and Discussion

People chose the more creditworthy company in 64.6% of the trials. Assuming a uniform prior on people's performance, the 99% highest-posterior density credible interval is $[62.5\%;66.6\%]$. We can thus be more than 99% confident that people's average performance is above 62.5% and conclude that they performed significantly better than chance ($p < 10^{-15}$). This is qualitatively consistent with feature-based strategy selection but inconsistent with context-free strategy selection; see Figure 2. Consistent with Rieskamp and Otto (2006) performed worse on non-compensatory trials than on compensatory trials. Thus the deviation from optimal performance is not due to saving mental effort by using the simpler TTB strategy when the more demanding WADD strategy would be required. To measure learning we regressed our participants' average performance on the trial number and a constant. We found that the increase in our participants' performance was not statistically significant (95% CI of the slope: $[-0.1\%, 0.2\%]$). The observation that participants nevertheless performed above chance, suggests that they entered the experiment with moderately high strategy selection skills. Alternatively, participants could have used a single, more complex strategy that works for both kinds of problems, but note that with a larger number of trials systematic changes in strategy use have been demonstrated in a similar task; see Figure 1. The absence of evidence for learning is not necessarily incompatible with previous findings, because Rieskamp and Otto (2006) found small changes in performance after 168 trials in a simple environment, whereas our participants performed only 30 trials and the environment was very complex. Future experiments will employ a larger number of trials and more reliable feedback to explore whether people learn in the mixed decision environment too. In conclusion, people's performance in homogeneous decision environments is consistent with feature-based and context-free strategy selection, but context-free strategy selection is insufficient to explain human performance in heterogeneous environments whereas feature-based strategy selection can account for it. Thus our results suggest that human strategy selection is feature-based.

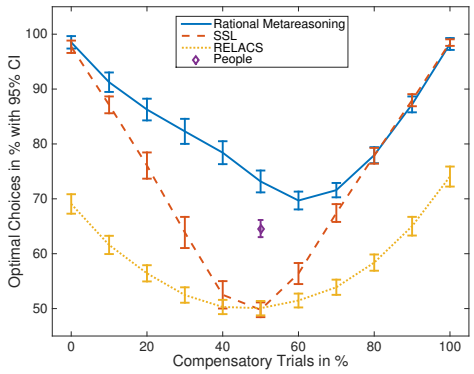


Figure 2: Rational metareasoning outperforms SSL and RELACS–especially when the environment is heterogeneous.

## Adaptive flexibility in strategy selection

People adapt their strategy not only to reoccurring situations, but they can also flexibly switch strategies as soon as the situation changes. This flexibility has been empirically demonstrated in decision making under risk: Payne et al. (1988) found that people adaptively switch decision strategies in the absence of feedback.

To determine whether rational metareasoning can account for this adaptive flexibility we simulated Experiment 1 from Payne et al. (1988). This experiment presented ten instances of each of four types of decision problems in random order. The four problem types were defined by the time constraint (15 seconds vs. none) and the dispersion of the outcomes' probabilities (low vs. high). In each decision problem participants chose between four gambles with four possible outcomes. The four gambles assigned different payoffs to the four outcomes but they shared the same outcome probabilities. The payoffs (range 0–999 cents) and their probabilities were stated numerically. To see an outcome probability or payoff participants had to click on the corresponding outcome. This allowed Payne et al. (1988) to infer which kind of strategies their participants were using. They measured the use of fast-and-frugal attribute-based heuristics, that is TTB and elimination-by-aspects (EBA; Tversky (1972)), by the proportion of time participants spend processing the options' payoffs for the most probable outcome. For the compensatory

expected value strategy WADD this proportion is only 25%, but for TTB and EBA it can be up to 100%. When the dispersion of outcome probabilities was high, people focused more on the most probable outcome. Time pressure also increased people's propensity for such selective and attribute-based processing; see Figure 3. Thus, people seem to use non-compensatory strategies such as TTB and EBA more frequently when time is limited or some outcomes are much more probable than others.

To simulate this experiment we applied rational metareasoning to choosing when to use which of the ten strategies considered by Payne et al. (1988). These strategies included WADD as well as fast-and-frugal heuristics such as TTB and EBA. To simulate the effect of the time limit we counted each strategy's elementary operations according to Johnson and Payne (1985), assumed that each of them takes one second, and returned the strategy's current best guess when it exceeded the limit as described by Payne et al. (1988). Rational metareasoning represented each risky-choice problem by five simple and easily computed features: the number of attributes, the number of options, the number of inputs per available computation, the highest outcome probability, and the difference between the highest and the lowest payoff. These features were manually chosen to capture highly predictive dimensions along which decision problems were varied by Payne et al. (1988). Our rational metareasoning model of strategy selection in risky choice learns to predict each strategy's relative reward

$$r_{\text{rel}}(s) = \frac{V(s(D),o)}{\max_c V(c,o)}, \tag{11}$$

where $s(D)$ is the gamble that strategy $s$ chooses in decision problem $D$, $V(c,o)$ is the payoff of choice $c$ if the outcome is $o$, and the denominator is the highest payoff the agent could have achieved given that the outcome was $o$. The priors on all coefficients in the score and execution time models of rational metareasoning were standard normal distributions.

We performed 1000 simulations of people's strategy choices in this experiment. In each simulation, we modeled people's prior learning experiences about risky choice strategies by applying rational metareasoning to ten randomly generated instances of each of the 144 types of decision problems considered by Payne et al. (1988). We then applied rational metareasoning with the learned model of the strategies' performance to a simulation of Experiment 1 from Payne et al. (1988). Since their participants received no feedback, our simulation assumed no learning during the experiment. Outcome distributions with low dispersion were generated by sampling unnormalized outcome probabilities independently from the standard uniform distribution and dividing them by their sum. Outcome distributions with high dispersion were generated by sampling the outcome probabilities sequentially such that the second largest probability was at most 25% of the largest one, the third largest probability was at most 25% of the second largest one, and so on.

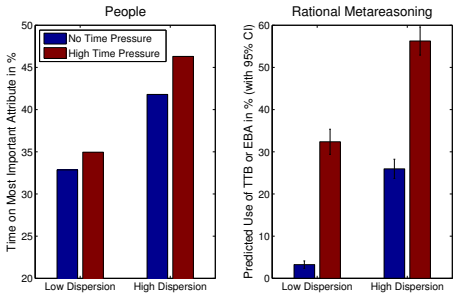


Figure 3: Rational metareasoning predicts the increase in selective attribute-based processing with dispersion and time pressure observed by Payne et al. (1988).

Rational metareasoning correctly predicted that time-pressure and probability dispersion increase people's propensity to use TTB or EBA; see Figure 3. SSL and RELACS, by contrast, predict that there should be no difference between the four conditions. This is because SSL and RELACS cannot learn to choose different strategies for different kinds of problems. Their strategy choices only change in response to reward or punishment but the experiment provided neither. In conclusion, rational metareasoning can account for adaptive flexibility in decision making under risk but SSL and RELACS cannot.

## General Discussion

We have proposed a rational solution to the strategy selection problem: feature-based strategy selection by rational metareasoning. We have previously shown that feature-based strategy selection can – but context-free strategy selection cannot – account for people's adaptive choice of sorting strategies (Lieder, Plunkett, et al., 2014). Here we have extended this conclusion to decision-strategies that operate on internal representations. The theoretical significance of rational metareasoning is twofold: First, it reconciles the two poles of the debate about human rationality by proposing that people learn to make rational use of fallible heuristics. Our theory formalizes the strategy selection principle of ecological rationality (Todd & Gigerenzer, 2012) and specifies the computational mechanisms of learning and cognitive control through which it could be realized. Second, strategy selection by rational metareasoning completes the resource-rational approach to cognitive modeling: the principle of resource-rationality can be used not only to derive heuristics (Lieder et al., 2013; Griffiths et al., 2015; Lieder, Hsu, & Griffiths, 2014) but also to predict when people should use which heuristic. Strategy selection by rational metareasoning thereby provides the conceptual glue necessary to build integrated theories out of our scattered bricks of knowledge about tens or hundreds of cognitive strategies.

If the brain had sufficient computational power and sophistication to perform near-optimal strategy selection, then why would it rely on simple heuristics to make economic decisions? The reason might be that real-life decisions are much

more complex than the decisions modeled above: The computational complexity of optimal decision making increases with the size of the decision problem but the computational complexity of strategy selection does not. Thus, as decision problems become more complex normative decision strategies take prohibitively long, but the time required for strategy selection remains the same. Therefore, rational metareasoning is not only tractable but also worthwhile in complex real-life decisions where the normative solution is intractable and applying the wrong heuristic can have dire consequences.

The exact conditions under which the benefits of strategy selection by rational metareasoning offset its cost will be determined in future work. In addition, rational metareasoning can be a used as a computational level theory (Marr, 1983) of metacognition–just like Bayesian inference is a used as computational level theory of inductive reasoning and learning.

In conclusion, strategy selection by rational metareasoning is a promising framework for cognitive modeling. It could, for instance, be used to explain paradoxical inconsistencies in risky choice by identifying why people use different heuristics in different contexts. Furthermore, our model of strategy selection learning can be applied to education and cognitive training: First, our model could be used to optimize problem sets for helping students learn when to apply which procedure (e.g. in algebra) rather than drilling them on one procedure at a time. Second, our model could be used to design cognitive training programs promoting adaptive flexibility in decision making and beyond. Future work will also explore learning the VOC of elementary information processing operations (Russell & Wefald, 1991) as a model of strategy discovery.

**Acknowledgments.** This work was supported by ONR MURI N00014-13-1-0341 and grant number N00014-13-1-0341 from the Office of Naval Research.

## References

Bröder, A. (2003). Decision making with the" adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *J. Exp. Psychol.-Learn. Mem. Cogn.*, *29*(4), 611.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological review*, *112*(4), 912–931.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.*, *14*(3), 119–130.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482.

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

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

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

Lieder, F., Griffiths, T. L., & Goodman, N. D. (2013). Burn-in, bias, and the rationality of anchoring. In P. Bartlett, F. C. N. Pereira, L. Bottou, C. J. C. Burges, & K. Q. Weinberger (Eds.), *Adv. neural inf. process. syst. 25*.

Lieder, F., Hsu, M., & Griffiths, T. L. (2014). The high availability of extreme events serves resource-rational decision-making. In *Proc. 36th ann. conf. cognitive science society*. Austin, TX: Cognitive Science Society.

Lieder, F., Plunkett, D., Hamrick, J. B., Russell, S. J., Hay, N., & Griffiths, T. (2014). Algorithm selection by rational metareasoning as a model of human strategy selection. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Weinberger (Eds.), *Adv. neural inf. process. syst. 27* (pp. 2870–2878). Curran Associates, Inc.

Marewski, J. N., & Link, D. (2014). Strategy selection: An introduction to the modeling challenge. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 39–59.

Marr, D. (1983). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman.

May, B. C., Korda, N., Lee, A., & Leslie, D. S. (2012). Optimistic Bayesian sampling in contextual-bandit problems. *J. Mach. Learn. Res.*, *13*.

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (2011). The recognition heuristic: a review of theory and tests. *Frontiers in psychology*, *2*.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *J. Exp. Psychol.-Learn. Mem. Cogn.*, *14*(3), 534.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *J. Exp. Psychol. Gen.*, *135*(2), 207–236.

Russell, S., & Wefald, E. (1991). Principles of metareasoning. *Artificial Intelligence*, *49*(1-3), 361–395.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*(5), 405–410.

Stanovich, K. E. (2009). *Decision making and rationality in the modern world*. Oxford University Press, USA.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 285–294.

Todd, P. M., & Gigerenzer, G. (2012). *Ecological rationality: Intelligence in the world*. New York: Oxford University Press.

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

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

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive science*, *38*(4), 599–637.