AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

https://doi.org/10.1037/rev0000330

Heuristics From Bounded Meta-Learned Inference

Marcel Binz[1, 2], Samuel J. Gershman[3, 4], Eric Schulz[2], and Dominik Endres[1]

[1] Department of Psychology, University of Marburg

[2] Computational Principles of Intelligence Lab, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

[3] Department of Psychology and Center for Brain Science, Harvard University

[4] MIT, Center for Brains, Minds, and Machines, Cambridge, Massachusetts, United States

Numerous researchers have put forward heuristics as models of human decision-making. However, where such heuristics come from is still a topic of ongoing debate. In this work, we propose a novel computational model that advances our understanding of heuristic decision-making by explaining how different heuristics are discovered and how they are selected. This model—called bounded meta-learned inference (BMI)—is based on the idea that people make environment-specific inferences about which strategies to use while being efficient in terms of how they use computational resources. We show that our approach discovers two previously suggested types of heuristics—one reason decision-making and equal weighting—in specific environments. Furthermore, the model provides clear and precise predictions about when each heuristic should be applied: Knowing the correct ranking of attributes leads to one reason decision-making, knowing the directions of the attributes leads to equal weighting, and not knowing about either leads to strategies that use weighted combinations of multiple attributes. In three empirical paired comparison studies with continuous features, we verify predictions of our theory and show that it captures several characteristics of human decision-making not explained by alternative theories.

Keywords: meta-learning, resource rationality, heuristics, strategy selection, strategy discovery

Imagine having to decide which of two movies you are going to watch tonight: Movie *A* versus Movie *B*. Movie *A* has a higher average rating on a website that you trust, while Movie *B* is directed by a known director and has previously won an Oscar for the best picture. From past experiences, you know that rating is the best indicator for a good movie. Whether the movie won an Oscar and who directed it is less important for how much you normally enjoy watching a movie. How do people make decisions like this?

The question of how people decide between two options is as fundamental as its answer is contentious. Indeed, even though we make countless such decisions every day, the underlying principles of these decisions are still debated in psychology (Todd & Gigerenzer, 2000), behavioral economics (Samuels et al., 2012), and neuroscience (Camerer et al., 2005). Traditionally, researchers have approached this problem by looking at how rational agents decide. From this ideal observer perspective (Geisler, 1989) it is assumed that people weigh different attributes of each option appropriately to combine information from all available sources. Psychologists were however quick to point out that rational decision-making can be too burdensome (Simon, 1990b; Tversky & Kahneman, 1974). Instead, they suggested that human decision-making may be based on a variety of heuristics, which are simple strategies that ignore part of the relevant information (Gigerenzer & Todd, 1999; Shah & Oppenheimer, 2008; Tversky & Kahneman, 1974).

Two common classes of heuristics are *one reason decision-making* (Gigerenzer & Goldstein, 1999) and *equal weighting* (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975). One reason decision-making heuristics are based on the idea that good reasoning often requires just a single piece of information (Marewski et al., 2010). Applying such a strategy to the initial example, you would only need to inspect the most important attribute: The movie rating. Based on this attribute, you decide to watch Movie *A* and ignore all other information about both movies. Equal weighting heuristics on the other hand completely abstain from differentiating between the attributes and instead tally all of them together to decide which option to choose. In our example, Movie *B* has two attributes in its favor, while Movie *A* only has one. Hence, you would decide to watch Movie *B* if your decision was based on an equal weighting heuristic.

Even though they are computationally simplistic strategies, heuristics can be surprisingly competitive in many real-world benchmarks (Czerlinski et al., 1999; Lichtenberg & Şimşek, 2017). This observation led different researchers to consider heuristics as

Marcel Binz https://orcid.org/0000-0001-8872-8386
Samuel J. Gershman https://orcid.org/0000-0002-6546-3298
Dominik Endres https://orcid.org/0000-0001-9756-9655

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 290878970-GRK 2271, Project 2. Eric Schulz is supported by the Max Planck Society and the Jacobs Foundation. Samuel J. Gershman is supported by the Center for Brains, Minds, and Machines (funded by NSF STC award CCF-1231216).

We did not publish our results in any form at an earlier point. A preliminary version of our model simulations was presented orally at the 1st GK Doctoral Symposium on Cognitive Science (Tübingen, 2020). The code to reproduce the results from this article is publicly available under https://github.com/marcelbinz/HeuristicsFromBMLI. A preprint is available under https://psyarxiv.com/5du2b.

The authors would like to thank Björn Meder, Lion Schulz, and Raphael Schween for comments on an earlier draft.

Correspondence concerning this article should be addressed to Marcel Binz, Computational Principles of Intelligence Lab, Max Planck Institute for Biological Cybernetics, Max-Planck-Ring 8, 72076 Tübingen, Germany. Email: marcel.binz@tue.mpg.de

ecologically rational strategies (Gigerenzer & Gaissmaier, 2011; Gigerenzer & Todd, 1999; Payne et al., 1993), implying that heuristics are strategies that are particularly well-suited for our complex and dynamic world. The ecological rationality of heuristics also makes it appealing to view them as models of human decision-making. Empirical studies attempting to show that people apply heuristics have however produced mixed evidence (Ayal & Hochman, 2009; Bröder, 2000; Bröder & Gaissmaier, 2007; Glöckner & Betsch, 2008; Hilbig, 2010, see also our later discussion on empirical results).

In this work, we suggest *bounded meta-learned inference* (BMI) as a novel computational theory for explaining how people make decisions. BMI discovers decision-making strategies through a resource-rational algorithm (Gershman et al., 2015; Lieder & Griffiths, 2019; Simon, 1990a) that has been adapted to an environment over time via meta-learning (Bengio et al., 1991; Schmidhuber et al., 1996; Thrun & Pratt, 1998). Like ideal observer models, BMI attempts to infer optimal decision-making strategies but does so while taking computational resources into account. Like heuristics, strategies inferred through BMI are tailored to a specific environment. However, unlike heuristics, the inductive biases of such strategies have been meta-learned through previous interactions with the environment instead of building them in by design.

Through a series of model simulations, we demonstrate that BMI discovers several previously suggested heuristics. Specifically, our results reveal three important classes of environments that lead to three different strategies. First, if the model knows the correct ranking of attributes but not their weights, then it learns a strategy that makes decisions based only on the attribute with the highest ranking, a form of one reason decision-making. Secondly, if the model knows that the direction of correlation between attributes and outcome is positive, then it learns a strategy that makes decisions based on equal weighting. Finally, if the model does not know either the ranking or the direction of attributes, it learns to use individual weights for each attribute. This analysis provides new insights into the mixed results of prior empirical work on heuristics because it makes precise predictions about if and when a specific heuristic should be used. We verify these predictions in three empirical paired comparison studies and show that the vast majority of participants apply heuristics whenever they are optimal strategies for the current environment after considering limited computational resources.

In summary, our work makes the following three main contributions:

1. We show that heuristics can emerge through BMI, thereby providing a normative justification for previously suggested heuristics.
2. We map out which features of an environment lead to which (heuristic) decision-making strategy, where knowing the correct ranking of attributes leads to one reason decision-making, knowing the directions of the attributes leads to equal weighting, and not knowing about either leads to strategies that use weighted combinations of multiple attributes.
3. We test these predictions empirically in three experiments and find strong evidence for our theory's predictions.

The remainder of the article is organized as follows: We first summarize the relevant literature on heuristic decision-making and introduce its general terminology. Thereafter, we present formal models corresponding to different hypotheses considered in our work. By running simulations on different environments, we generate several predictions of our theory, which we empirically test in three new decision-making experiments. Finally, we discuss our results and connect our theory to related ideas.

Past Research on Heuristic Decision-Making

There has been an extensive amount of past research on heuristic decision-making. In this section, we describe common heuristics, summarize empirical and theoretical results regarding their performance, and review prior studies with a focus on the evidence they provide for heuristic decision-making in humans.

Heuristics Toolbox

Even though a mathematically precise definition of what constitutes a heuristic is still a topic of ongoing debates (Chater et al., 2003; van Rooij et al., 2012), here we adopt the following definition put forward by Gigerenzer and Gaissmaier (2011): "A heuristic is a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods." The collection of different heuristics is often thought of as an adaptive toolbox from which appropriate decision-making strategies can be selected as required (Gigerenzer & Selten, 2002). We are primarily interested in heuristics that can be applied to paired comparison tasks like the aforementioned movie example (e.g., Martignon & Hoffrage, 2002). In such tasks, a decision-making agent is asked to judge which of two options is superior on an unobserved criterion. To aid the decision-making process, the agent observes multiple attributes of both options, also known as cues or features in the decision-making literature. Most heuristics developed for the paired comparison setting make use of binary features that indicate whether an attribute is present or not.[1]

Many decision-making strategies are built around the concept of feature validity (Todd & Dieckmann, 2005). The validity of a binary feature is the rate at which it allows the agent to make correct predictions given that the feature is present in one option but not the other (Lee & Cummins, 2004). For example, the validity of being directed by a known director for predicting whether you like a movie could be 0.8, indicating that you would enjoy a movie that is directed by someone you know over someone you do not know in 80% of the cases. In general, decision-making strategies for paired comparison tasks can be divided into two classes: compensatory and noncompensatory strategies. A strategy is compensatory whenever it integrates information from multiple features, whereas it is noncompensatory when a feature cannot be outweighed by any combination of less important features (Rieskamp & Hoffrage, 1999).

A prominent subclass of compensatory strategies are linear-additive strategies. These strategies compute a weighted sum of features for each option and decide on the option with the largest sum.

[1] Note that nonbinary features, like average movie ratings, can always be dichotomized at a loss of information. In past studies, this has been frequently done by setting values which were less than the median to 0 and otherwise to 1.

They are typically considered the normative standard in the decision-making literature (Payne et al., 1988). This argument can be made precise if one weights features by the log-odds of their validities. The resulting strategy corresponds to an algorithm known as naive Bayes, which is optimal under the assumption that features are conditionally independent given the criterion (Katsikopoulos & Martignon, 2006; Lee & Cummins, 2004). The weighted additive (WADD) strategy is another popular example of a linear-additive strategy, which weights features directly by their validities. In contrast to naive Bayes, however, it is not possible to interpret WADD as an optimal strategy.

Heuristics are typically much simpler than WADD or naive Bayes. Equal weighting heuristics, for example, are compensatory, yet simple, decision-making strategies. They do not distinguish between how features are weighted and instead use an identical weighting for all features. The process itself can be realized by tallying features of both options together and selecting the one with the larger sum (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975).

The prime example for a noncompensatory strategy is the take-the-best (TTB) heuristic (Gigerenzer & Goldstein, 1996). TTB belongs to the family of one reason decision-making heuristics. It assumes a ranking of features based on their validities and inspects features in decreasing order until a feature that discriminates between both options is reached. The final decision is based on the validity of that feature alone, ignoring all other information. If a ranking of features is not a priori accessible, then it can either be estimated from observations or a random ranking can be used. A TTB strategy using a random ranking of features is referred to as the Minimalist heuristic (Gigerenzer & Goldstein, 1996).

Ecological Rationality

To seriously consider heuristics as a model of human decision-making, they should—at the very least—be able to solve the kind of decision-making problems that people typically encounter. Prior work demonstrated that heuristics do not only match the performance of more complex linear-additive models but even exceed them on such problems. This finding is also referred to as the less-is-more effect (Gigerenzer & Todd, 1999). Czerlinski et al. (1999), for example, compared different heuristics against logistic regression on 20 real-world decision-making problems and found that averaged over all tasks TTB and logistic regression performed equally well. Chater et al. (2003) extended this analysis to additionally include a feed-forward neural network, two exemplar-based models, and a decision-tree algorithm. They concluded that the less-is-more effect is most prevalent when only limited data is available. Later on, it was highlighted that, although earlier work fitted models on a limited training set, it evaluated them on the entire data-set (training and test set). It turned out that, when only out-of-sample predictions were considered, TTB even exceeded all competing models in terms of performance (Brighton, 2006; Gigerenzer & Brighton, 2009; Katsikopoulos et al., 2010). More recently, Lichtenberg and Şimşek (2017) have shown that the less-is-more effect also extends to situations in which one has to make predictions about a continuous outcome.

The discovery of the less-is-more effect caused researchers to ask themselves, why do heuristics perform so well? Eventually, this cumulated in several conditions that allow us to make claims about the performance of a heuristic based on the structure of the task it is applied to (Katsikopoulos, 2011). In the case of binary features, it has been shown that decisions made by TTB are identical to those of a linear-additive model if the true feature weights of the task are noncompensatory, that is, when a more important feature cannot be overruled by any combination of less important features (Martignon & Hoffrage, 1999, 2002). A similar result was obtained by Katsikopoulos and Martignon (2006) under the assumption that features are conditionally independent given the criterion. Baucells et al. (2008) described a different task structure known as cumulative dominance that causes both TTB and equal weighting to achieve maximum performance across all strategies. An option cumulatively dominates another option—under the assumption that features are ordered according to their importance—if all of its cumulative sums of features are larger than the ones of the alternative option. Şimşek (2013) demonstrated that both noncompensatoriness and cumulative dominance are relatively common in many real-world decision-making problems and therefore provided a justification for the use of heuristics.

A related line of research has argued that heuristics work well because they involve fewer parameters, and are hence easier to learn based on limited or noisy observations. In this context, Hogarth and Karelaia (2005, 2006, 2007) derived several analytical conditions under which different heuristics—like TTB and equal weighting—achieve superior performance compared to a linear-additive model whose weights are estimated using maximum likelihood estimation. In particular, they found that both heuristics perform well when the number of observations used for estimation is small compared to the number of features. They additionally demonstrated that TTB tends to perform well when the variability of validities between features is high, while equal weighting performs well when the variability of validities is low. Gigerenzer and Brighton (2009) approached the less-is-more effect from the perspective of the bias-variance trade-off, which states that the expected generalization error of a model can be decomposed into the sum of a bias and a variance component. A model with high bias fails to capture regularities in the data, while a model with high variance is sensitive to small fluctuations in the sample. Gigerenzer and Brighton (2009) argued that if observations are sparse or noisy the variance component will typically dominate and that heuristics achieve superior performance in such conditions because they keep this component within acceptable limits. Şimşek and Buckmann (2015) provided additional support for this theory by showing that building blocks of different heuristics can be learned efficiently with just a few training samples. Finally, Parpart et al. (2018) argued that heuristics can also emerge from Bayesian inference in the limit of infinitely strong priors. Based on this insight, they identified priors that correspond to both TTB and equal weighting. Their work indicated that heuristics perform well because they implement strong priors that approximate the actual structure of the environment.

Thus far, we have discussed environmental conditions that favor heuristic decision-making. There exists, however, a complementary justification for why people should use heuristics: They simply involve less complicated computations (Payne et al., 1988, 1993). Shah and Oppenheimer (2008) have advocated for the study of heuristics in the light of the accuracy-effort trade-off. From their point of view, heuristics are interpreted as strategies that save effort at the cost of a potentially reduced accuracy. Resource-rational

theories of decision-making take this point of view one step further (Bhui et al., 2021). Instead of only asking how to save computational resources, resource-rational models identify how to spend a limited amount of resources optimally in order to maximize accuracy. Lieder et al. (2017) applied the framework of rational metareasoning to construct resource-rational decision-making strategies. They showed that TTB can be considered rational if execution time is limited. We currently know very little about whether common heuristics can be interpreted as strategies that make optimal use of limited computational resources beyond the results of Lieder et al. (2017).

Empirical Studies

The observation that heuristics are computationally efficient and ecologically rational strategies is often used to justify them as models of human decision-making (Todd & Gigerenzer, 2007). However, to truly establish that people use heuristics, proving good performance in simulation and theory is not sufficient; it also requires empirical evidence. Many studies have attempted to find such evidence, yet no consensus for or against heuristics has been reached. Here, we provide an overview of these studies and attempt to connect their findings. While we focus on studies in the paired comparison setting, we also included a few notable exceptions with more than two choice alternatives.

Evidence for Heuristics

Let us first consider studies that provided evidence for heuristics. The majority of such evidence comes from studies in which it was costly to access information about feature values. The Mouselab paradigm is a process-tracing approach to decision-making, which requires participants to click or hover over a specific feature to reveal its value. In studies making use of the paradigm, Rieskamp and Otto (2006) showed that people's selection of strategies depended on the environment they interacted with. Participants in their study had initial preferences for linear-additive strategies, but then slowly adopted TTB in a noncompensatory environment and WADD in a compensatory environment. Mata et al. (2007) confirmed this general result in a study with participants from different age groups. They found that across all age groups, participants looked up less information in an environment with unequal validities compared to one with equal validities. They also concluded that older adults tend to select simpler strategies, like TTB, more frequently than their younger counterparts. Mata et al. (2010) reinforced the hypothesis that older people tend to apply simpler strategies. In particular, they demonstrated that older people were more likely to apply equal weighting—instead of WADD—in a compensatory environment. In a similar paradigm, Wichary et al. (2016) demonstrated that placing participants under emotional stress caused them to search for less information and to apply simpler strategies.

Another way to promote the use of heuristics is to require a monetary fee to reveal features. In several experiments with monetary fees, Bröder (2000) produced evidence in favor of one reason decision-making heuristics. In his experiments, more participants were classified as TTB users in a high-cost condition compared to a low-cost condition. In another study, Bröder (2003) manipulated the payoff structure of the environment while keeping the nominal cost for obtaining information constant, that is, he considered environments in which it was advantageous to gather more information and those in which it was not. He found that most participants applied TTB in a noncompensatory environment, whereas they applied a linear-additive strategy when information was more valuable. In the latter condition, he also found that the percentage of non-TTB choices did not increase over time, suggesting that "a compensatory strategy may be something like a default strategy." Dieckmann and Rieskamp (2007) also observed that TTB predicted more decisions in environments with monetary costs and furthermore demonstrated that participants applied TTB more often when the redundancy between features was high.

It has also been argued that people rely more on heuristic decision-making when feature values are not readily available but have to be retrieved from memory instead. In multiple experiments with memory-based retrieval, Bröder et al. demonstrated that participants became more consistent with TTB when features had to be retrieved from memory (Bröder & Gaissmaier, 2007; Bröder & Schiffer, 2003, 2006). Bröder and Schiffer (2003), for example, classified 72% of participants as TTB users when they were under high working memory load, but only 56% when they were not. Persson and Rieskamp (2009) used a similar paradigm but required participants to learn about the interaction between features and the criterion based on feedback. They found that most participants applied TTB when feedback indicated which option was better on the unobserved criterion. However, when direct feedback about criterion values was provided, most people applied a linear-additive strategy instead. In addition, they also included an exemplar-based approach in their analysis but found little evidence for such a strategy.

What about situations in which information is freely available? There exists overall only limited evidence suggesting that people apply heuristics in such cases. Bergert and Nosofsky (2007) were among the few who provided support for the idea that people rely on heuristics even when information is free. In their study, participants exhibited noncompensatory decision-making patterns, assigning over half of the total weight to a single feature. They further strengthened their claim using a novel reaction time method that allowed them to disentangle predictions of different strategies.

Evidence Against Heuristics

In general, it seems that increasing the costs for utilizing information can make human decision-making more consistent with heuristics. However, even under such supposedly favorable conditions, it is still disputed whether people use heuristics or if they rely on more complex strategies instead. Newell et al. (2003) demonstrated that even with large monetary costs and other conditions favoring one reason decision-making heuristics, not many participants acted fully in accordance with TTB. When reanalyzing the data of Rieskamp and Otto (2006), Scheibehenne et al. (2013) found that people in noncompensatory environments were better described through a mixture of TTB and WADD, indicating a general preference for linear-additive strategies. van Ravenzwaaij et al. (2014) showed that hierarchical models accounting for both search order and termination provided a better explanation for participants' choices than TTB and WADD alone. Similarly, Söllner and Bröder (2016) found that people tend to adjust how long then search for information based on the evidence that they have accumulated so far. They concluded that this observation is in line with evidence

accumulation models (Lee & Cummins, 2004), but not with heuristics like TTB.

We also have to be cautious to not misinterpret evidence for heuristics in process tracing studies as a general inability to apply more complex strategies when information search is not constrained by the experimental paradigm. In this context, Glöckner and Betsch (2008) argued that process-tracing studies are likely to underestimate the cognitive capacity of participants, as they hinder the activation of automatic decision-making processes. They verified this claim by demonstrating that participants were generally able to combine information from multiple features extremely quickly when the acquisition of information was not constrained. Further studies with freely accessible information provided similar results (Bröder, 2000; Heck et al., 2017; Lee & Cummins, 2004; Parpart et al., 2018), always concluding that few participants made decisions consistent with TTB and that their choices were, in general, better described through linear-additive strategies. Finally, Newell and Lee (2011) highlighted large interindividual differences and presented a sequential sampling model that provided better fits than TTB, WADD, and a strategy selection model across all participants.

Heuristics in Related Research Areas

There are also a number of research areas that use experimental paradigms similar to paired comparison studies, which have produced mixed evidence on whether people rely on heuristic decision-making or not. In probabilistic category learning (Ashby & Maddox, 2005), participants are asked to classify objects into one of usually two categories. Thus, similar to paired comparison tasks, participants learn a mapping between features and a binary outcome. Juslin, Olsson, et al. (2003) noted that the category learning literature emphasizes exemplar models, which is in contrast to the linear-additive models studied in the decision-making literature. Based on this observation, they investigated which factors modulate a shift from exemplar models to linear-additive cue-integration models. In a similar setting, von Helversen et al. (2013) demonstrated that participants switched from an exemplar-based model to a linear-additive cue-integration model once information about the direction of features was available. However, both of these lines of work did not examine the role of heuristics in the context of category learning. In a later study, Juslin, Jones, et al. (2003) did consider the possibility for one reason decision-making heuristics but found little evidence for such strategies, even after introducing additional time pressure.

Another closely connected paradigm with a long history on its own is multiple-cue probability learning (MCPL, Brehmer, 1979; Gluck & Bower, 1988; Hammond, 1955). In MCPL, people have to learn about a probabilistic relationship between an object described by multiple features and an outcome. A popular instance of MCPL is given by the weather prediction task. Here, participants are presented with a multidimensional stimulus taking the form of tarot cards and learn based on feedback whether given patterns lead to sunny or rainy weather. Gluck et al. (2002) conjectured that people approach this task using three different strategies: (a) an optimal strategy, which learns about all available features, (b) a one reason decision-making heuristic, in which decisions are based on a single feature, and (c) a singleton heuristic, which learns only about the patterns that have a single feature present. In two studies, they found that a majority of participants (85% across both studies) was overall best fit by the singleton heuristic. However, as more data were observed participants either switched toward the one reason decision-making heuristic in a more challenging experiment or the optimal multicue strategy in an easier experiment. In contrast, Lagnado et al. (2006) concluded that a vast majority of participants was best described by a strategy that integrated information from all features (86% across three studies). Newell et al. (2007) reported similar results, with the additional observation that people switched toward a more simplistic singleton heuristic if they were put under working memory load. Finally, it is worth pointing out that equal weighting also received some attention in the MCPL literature: When participants were provided with directional information about features, they switched from a multicue strategy toward equal weighting (Newell et al., 2009). In the context of this article, this is an interesting observation, because—as we will show later on—it is exactly what our model predicts.

Summary

To summarize, many prior studies attempted to produce evidence for one reason decision-making strategies like TTB, while focusing less on other heuristics such as equal weighting. They concluded that such strategies were indeed more apparent when it was costly to access information about feature values, either in terms of time, money, or memory. However, when features were freely accessible, evidence for heuristics in human decision-making remained rare. Where does that leave us? We argue, following earlier work of Lieder et al. (2017), Payne et al. (1993), and Shah and Oppenheimer (2008), that examining which strategies are rational after taking limited computational resources into consideration can help us to understand why and when people use heuristics. In the next section, we formalize this idea and present a novel modeling framework that allows us to determine which strategy is resource-rational for a particular environment. We then show—with the help of this framework—that previously suggested heuristics can be interpreted as resource-rational solutions to paired comparison tasks when additional information about the ranking or the direction of features is available. In our experimental studies, we indeed find that people use their computational resources efficiently, and apply heuristics when such information is available. This result is amongst the first to show that people rely heavily on heuristics even in the absence of time, money, or memory constraints. However, when no side information is available, we find that it is resource-rational to make decisions using weighted combinations of features. We confirmed empirically that under such conditions, people do not rely on heuristics and instead apply linear-additive strategies. This result is in agreement with the majority of reviewed studies with freely accessible feature values.

Computational Models

In this section, we show how recent advances in meta-learning can be used to construct environment-specific decision-making algorithms that make optimal use of limited computational resources. Having access to such an algorithm does allow us to predict if and when people should rely on heuristic decision-making strategies, assuming that they use available computational resources efficiently. To test this conjecture, we also introduce several other computational models of decision-making in paired comparison tasks. First, we will outline the assumptions about the structure of

the problem to be solved and define a corresponding ideal observer model. Then, we introduce probabilistic variants of two popular heuristics. Both heuristics are considerably simpler than an ideal observer model regarding their use of available information. Finally, we describe how we obtain resource-rational algorithms that are adapted to a particular environment.

The decision-making problems we focus on in this article are paired comparison tasks with continuous features. In a paired comparison task an agent—either human or machine—has to decide which of two options with feature vectors $\mathbf{x}_{A,B} \in \mathbb{R}^d$ has the higher value on an unobserved criterion $y_{A,B}$. In our movie example, the feature vector contains information about whether the movie has won an Oscar, its average rating on a reviewing website, and so on, while the unobserved criterion corresponds to your personal rating of the movie (i.e., how much you would like the movie if you watched it). We consider the setting where data arrive sequentially, that is, one at a time, and with feedback that indicates which option has the higher criterion value. Let $\mathbf{x}_{A,t}$ and $\mathbf{x}_{B,t}$ denote the observed features at time-step t and let c_t be a binary variable that takes the value of 1 if option A has the higher criterion value and 0 otherwise. For each time-step, the agent first observes both options, then makes a prediction about c_t, and subsequently receives feedback about which option had the higher criterion value. Note that learning in this setting is always based on feedback in the form of c_t, and that the criteria $y_{A,t}$ and $y_{B,t}$ are never observed directly.

In contrast to most prior work, we investigate paired comparison tasks with continuous features. In many real-world scenarios, features are naturally described through continuous values and thus we believe that the restriction to binary features oversimplifies a characteristic present in many of the problems people typically solve. Moving to continuous features also facilitates statistical analysis as fewer trials are needed to observe expected effects. For example, it would require over four times more trials to distinguish an ideal observer model from a single cue heuristic in environments with dichotomized features instead of continuous ones (see Appendix A for further details).

Ideal Observer

Ideal observer models are designed to provide a theoretical upper bound on performance in a specific task. In the following, we construct an ideal observer model for paired comparison tasks. For this, we assume that there exists an underlying linear relationship between features and the criterion:

$$\begin{aligned} y_A &= \mathbf{w}^T \mathbf{x}_A + \epsilon_A \\ y_B &= \mathbf{w}^T \mathbf{x}_B + \epsilon_B, \end{aligned} \tag{1}$$

with feature weights $\mathbf{w} \in \mathbb{R}^d$ and independent, additive noise $\epsilon_{A,B} \sim \mathcal{N}(0, \sigma^2)$. Based on this assumption, we can express the probability, that option A has a higher criterion value than option B as:

$$\begin{aligned} p(Y_{A,t} > Y_{B,t} | \mathbf{x}_{A,t}, \mathbf{x}_{B,t}, \mathbf{w}, m = \text{IO}) &= p(C_t = 1 | \mathbf{x}_t, \mathbf{w}, m = \text{IO}) \\ &= \mathbf{\Phi}\left(\frac{\mathbf{w}^T \mathbf{x}_t}{\sqrt{2}\sigma}\right), \end{aligned} \tag{2}$$

where $\mathbf{\Phi}$ is the cumulative distribution function of a standard normal distribution. For ease of notation, we have denoted the difference between feature vectors as $\mathbf{x}_t = \mathbf{x}_{A,t} - \mathbf{x}_{B,t}$ and used the binary variable C_t to indicate which of the two options has a higher criterion value.

Equation 2 is known in the statistics and machine learning literature as the probit regression model. The probit regression model makes it clear that an ideal observer should represent the probability that one option is better than the other using a weighted sum of differences between features of the options. Hence, the ideal observer model is a compensatory decision-making strategy.[2]

Parameter Estimation

Equation 2 provides an ideal observer model under the assumption that the underlying feature weights $\mathbf{w}$ are known. However, we assume that the weights are not provided in advance to the decision-maker. Thus, the agent has to infer them based on past observations. An ideal observer should apply Bayesian inference to infer unobserved parameters from data in a normative manner. In our setting, we estimate unobserved parameters by applying Bayesian inference sequentially. Exact inference is not possible under the above assumptions and thus we resort to a variational approximation (Jordan et al., 1999). We approximate the true posterior with a normal distribution $q(\mathbf{w}; \lambda_t) = \mathcal{N}(\mathbf{w}; \mu_t, \mathbf{\Psi}_t)$ and optimize its parameters $\lambda_t = (\mu_t, \mathbf{\Psi}_t)$ through gradient ascent on the evidence lower bound:

$$\begin{aligned} \mathcal{L}(\lambda_t) = \mathbb{E}_{\mathbf{w} \sim q(\mathbf{w}; \lambda_t)}[\log p(C_t = c_t | \mathbf{x}_t, \mathbf{w})] \\ - \text{KL}[q(\mathbf{w}; \lambda_t) || q(\mathbf{w}; \lambda_{t-1})], \end{aligned} \tag{3}$$

where $q(\mathbf{w}; \lambda_0)$ corresponds to an initial prior distribution. This kind of approximation is equivalent to exact inference when the true posterior is within the considered variational family. We provide further details on how Equation 3 is optimized in Appendix B.

To make predictions, we average over all plausible parameter values given by the variational distribution. The resulting predictive distribution can be expressed in closed form:

$$\begin{aligned} p(C_{t+1} = 1 | \mathbf{x}_{t+1}, \lambda_t) &= \int p(C_{t+1} = 1 | \mathbf{x}_{t+1}, \mathbf{w}) q(\mathbf{w}; \lambda_t) d\mathbf{w} \\ &= \mathbf{\Phi}\left(\frac{\mu_t^T \mathbf{x}_{t+1}}{\sqrt{2\sigma^2 + \mathbf{x}_{t+1}^T \mathbf{\Psi}_t \mathbf{x}_{t+1}}}\right). \end{aligned} \tag{4}$$

We assume, throughout this article, that features weights are sampled from a standard normal distribution at the beginning of each task and held constant over its entire duration, which implies that an ideal observer should use a prior in form of a standard normal distribution, that is, $q(\mathbf{w}; \lambda_0) = \mathcal{N}(\mathbf{w}; \mathbf{0}, \mathbf{I})$.

Heuristics

The two heuristics we consider in our analysis belong to the categories of one reason decision-making and equal weighting.

[2] Note that alternatively, it would have also been possible to assume that the noise term follows an extreme-value distribution, which would result in a logistic regression model. We have decided on the probit model instead because it allows us to compute predictive posterior distributions in closed-form.

In contrast to traditional heuristics, like TTB, they are probabilistic decision-making strategies for tasks with continuous features. Both are obtained through modification of the ideal observer model, such that either less information is required to make a decision or that information is combined in a simpler way.

One Reason Decision-Making

In our implementation of one reason decision-making, we modify Equation 2 and replace it with a model that only takes a single feature $\mathbf{x}_t^*$ into account:

$$p(C_t = 1|\mathbf{x}_t, w, m = \text{SC}) = \mathbf{\Phi}\left(\frac{w \cdot \mathbf{x}_t^*}{\sqrt{2}\sigma}\right). \quad (5)$$

We refer to the resulting strategy (Equation 5) as single cue heuristic. If a ranking of features is available, decisions are based on the most predictive feature, otherwise, we select the feature that performed best on the data so far. In contrast to TTB, the single cue heuristic does not involve a sequential search over features. However, we assume that features take continuous values, and hence search is not required as a feature nearly always discriminates between options (Luan et al., 2014). We have also experimented with a semilexicographic heuristic that uses a threshold parameter for deciding whether to consider a feature, but have not found it to explain the empirical data better than the simpler single cue heuristic. Thus, we decided to focus our analysis on the simpler implementation that always uses the first or best feature.

Equal Weighting

In our probabilistic version of equal weighting, we replace Equation 2 with a model that has a single, tied weight for all features:

$$p(C_t = 1|\mathbf{x}_t, w, m = \text{EW}) = \mathbf{\Phi}\left(\frac{w \cdot \sum_{i=1}^{d} \mathbf{x}_{t,i}}{\sqrt{2}\sigma}\right). \quad (6)$$

For $w > 0$, this equal weighting heuristic (Equation 6) probabilistically selects the option with the larger sum of features. For $w < 0$, it becomes more likely to select the option with the smaller sum (a negative weight is appropriate if most features have negative correlations with the criterion). Note that the ideal observer model contains as many free parameters as there are observed features, while both heuristics have only a single free parameter regardless of how many features are observed.

Bounded Meta-Learned Inference

Finally, we present BMI as a novel theory for human decision-making. BMI combines two equally important ideas: meta-learning and resource rationality. We are going to introduce them one after the other. First, we describe how meta-learning can produce decision-making algorithms that infer the optimal strategy for a particular environment, resulting in a variant without resource limitations called meta-learned inference (MI). Then, we show how MI can be extended to BMI by additionally taking limited computational resources into account. BMI may therefore be described as a resource-rational decision-making algorithm that has been adapted to an environment over time via meta-learning.

Meta-Learned Inference

Meta-learning (Bengio et al., 1991; Schmidhuber et al., 1996; Thrun & Pratt, 1998), also known as learning to learn, is a machine learning approach to devise learning systems that can rapidly adapt to new problems. In our work, we will use meta-learning as a purely methodological tool for constructing resource-rational decision-making algorithms, that is, we do not attempt to study the process of meta-learning itself but are only interested in its outcome.

The main idea of our approach is simple: Instead of using Bayesian (or variational) inference to infer posterior distributions over probit regression weights, we train a recurrent neural network to do this inference. In time-step $t + 1$, the network processes the previous feature vector $\mathbf{x}_t$ together with its corresponding target c_t, combines this information with its hidden state, and based on this estimates the parameters of the posterior distribution $\lambda_t = \{\mu_t, \mathbf{\Psi}_t\}$. Finally, it combines the estimated weights with the feature vector $\mathbf{x}_{t+1}$ as described in Equation 4 to obtain the predictive posterior distribution $p(C_{t+1} = 1|\mathbf{x}_{t+1}, \lambda_t, \mathbf{\Theta})$. Figure 1 illustrates graphically how the network processes a sequence of observations.

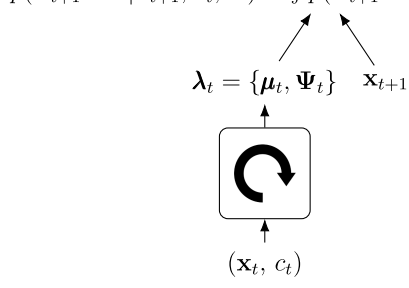
Initially, the recurrent network maps a sequence of previously observed feature-target pairs to a random posterior distribution over weights. During meta-learning, the system is then trained in an end-to-end manner to infer statistically optimal predictive posterior distributions for a distribution over tasks $p(\mathbf{x}_{1:T}, c_{1:T})$. We also refer to this distribution over tasks as the *environment*. In probabilistic terms, we can infer statistically optimal predictive posterior distributions by minimizing their negative log-probabilities on tasks sampled from the environment:

$$\mathcal{L}(\mathbf{\Theta}) = \mathbb{E}_{p(\mathbf{x}_{1:T}, c_{1:T})}\left[\sum_{t=0}^{T-1} -\log p(C_{t+1} = c_{t+1}|\mathbf{x}_{t+1}, \lambda_t, \mathbf{\Theta})\right], \quad (7)$$

where $\mathbf{\Theta}$ are parameters of the recurrent network, which we refer to as meta-parameters in order to distinguish them from the probit regression weights of Equation 2.

Equation 7 is optimized until convergence using standard optimization techniques. Through repeated encounters with the environment, the model is able to adapt to the properties of that specific

Figure 1
Graphical Depiction of MI/BMI



Note. MI = meta-learned inference; BMI = bounded meta-learned inference. The recurrent neural network sequentially processes examples from a given task. Through its recurrent activations it combines information from all previous feature-target pairs to compute a distribution over weights, which is then combined with the next input to obtain the predictive distribution.

environment. Once meta-learning has finished, the recurrent network acts as a free-standing decision-making algorithm without requiring any further optimization. Instead, adaptation to new tasks is simply implemented in form of forward passes through the network: We provide the network with a sequence of feature-target examples and an input that we want to query, and the network provides us with optimal predictive posterior distributions for that sequence of observations. We refer to the decision-making algorithm that is implemented by the forward dynamics of the recurrent network as MI. It has been shown in previous work that this meta-learning approach leads to the emergence of an algorithm that approximately simulates Bayesian inference (Mikulik et al., 2020; Ortega et al., 2019; Rabinowitz, 2019). Thus, MI will implement an algorithm similar to our ideal observer model, assuming that both of them make identical assumptions about the environment.

Resource Rationality

BMI is an extension to MI that additionally takes limited computational resources into consideration. More specifically, BMI controls for how many bits are required to implement the emerging decision-making algorithm, which is also referred to as its description length. From a psychological perspective, this may be interpreted as a cost for *storing* the algorithm that infers what decision-making strategies to apply. We will examine how this type of computational cost relates to other costs commonly used in cognitive science in the General Discussion.

How can this be formalized mathematically? First, we have to recognize that controlling the description length of meta-parameters is equivalent to controlling the description length of the emerging decision-making algorithm. This is the case because the emerging decision-making algorithm is fully specified through the meta-parameters. If we then represent the meta-parameters using a distribution over their plausible values $q(\boldsymbol{\Theta}; \boldsymbol{\Lambda})$, it is possible to interpret the *Kullback–Leibler* (KL) divergence between $q(\boldsymbol{\Theta}; \boldsymbol{\Lambda})$ and a prior $p(\boldsymbol{\Theta})$ as a measure of the meta-parameters' description length (Grünwald & Grunwald, 2007; Hinton & Van Camp, 1993).[3] In order to find the optimal balance between high performance with low computational complexity, BMI simply adds a β-weighted KL-term to the MI objective from Equation 7:

$$\mathcal{L}(\boldsymbol{\Lambda}) = \underbrace{\mathbb{E}_{p(\mathbf{x}_{1:T}, c_{1:T})}\left[\mathbb{E}_{q(\boldsymbol{\Theta};\boldsymbol{\Lambda})}\left[\sum_{t=0}^{T-1} -\log p(C_{t+1} = c_{t+1} | \mathbf{x}_{t+1}, \lambda_t, \boldsymbol{\Theta})\right]\right]}_{\text{performance}} + \beta \underbrace{\mathrm{KL}[q(\boldsymbol{\Theta}; \boldsymbol{\Lambda}) || p(\Theta)]}_{\text{description length}}. \quad (8)$$

In our later model comparisons, we treat β as a free parameter that is fitted to empirical data. For $\beta = 0$, meta-learning with Equation 8 is equivalent to MI. For $\beta > 0$, we get a family of decision-making algorithms that optimally trade-off performance for a shorter description length. In this article, we focus on the information-theoretic interpretation of Equation 8. For completeness, it should be noted that several authors have suggested an alternative interpretation that appeals to Probably Approximately Correct-Bayesian theory (McAllester, 2013), both in the context of traditional supervised learning (Achille & Soatto, 2018, 2019) and meta-learning (Yin et al., 2020).

We additionally use a sparsity-inducing prior (Kingma et al., 2015; Molchanov et al., 2017; Tipping, 2001), which means that under large resource limitations only networks with few nonzero meta-parameters remain. Thus, resulting algorithms are simple in terms of their description length and in terms of the number of remaining meta-parameters. Figure 2 schematically contrasts two networks obtained from optimization with low and high resource limitations. In Appendix C we provide a full specification of the network architecture, meta-learning procedure, and choice of prior.

It also seems sensible to ask: What decision-making strategies can BMI infer? Both the single cue heuristic and equal weighting are subsets of the space of all possible weight vectors that can be inferred. Equal weighting heuristics correspond to uniform vectors (e.g., [1, 1, 1]), while single cue heuristics can be expressed through a vector with a single nonzero entry (e.g., [1, 0, 0, 0]). BMI could thus—in principle—discover the two heuristics and select between them whenever appropriate. MI (or equivalently BMI with $\beta = 0$) approximately simulates an ideal observer, and hence we expect it to infer strategies that use independent and nonzero weights for all features. However, as we decrease the description length of the emerging decision-making algorithm, we expect it to infer simpler strategies like the single cue or equal weighting heuristic. Importantly, which strategy BMI infers, and whether it corresponds to a particular heuristic or not, does not only depend on its complexity but also on the distribution over tasks that was used for meta-learning.

Model Summary

Let us summarize all outlined models again and contrast the assumptions they make. The ideal observer model assumes that everything about the structure of the decision-making environment is known. In particular, it knows about the linear-Gaussian relationship. With this knowledge, it is able to compute the optimal solution by combining information from all features through weighted sums. Heuristics, like the single cue strategy and equal weighting, assume that computing weighted sums is too burdensome and instead bet on simpler ways for making decisions. The single cue heuristic only inspects a single feature, while the equal weighting heuristic sums up all features without weighting them. BMI does not know anything about the structure of the environment explicitly. Instead, it has acquired a resource-rational algorithm to infer decision-making strategies through repeated encounters with an environment. Thus, BMI can exploit characteristics present in that specific environment, while also being efficient in terms of computational resources.

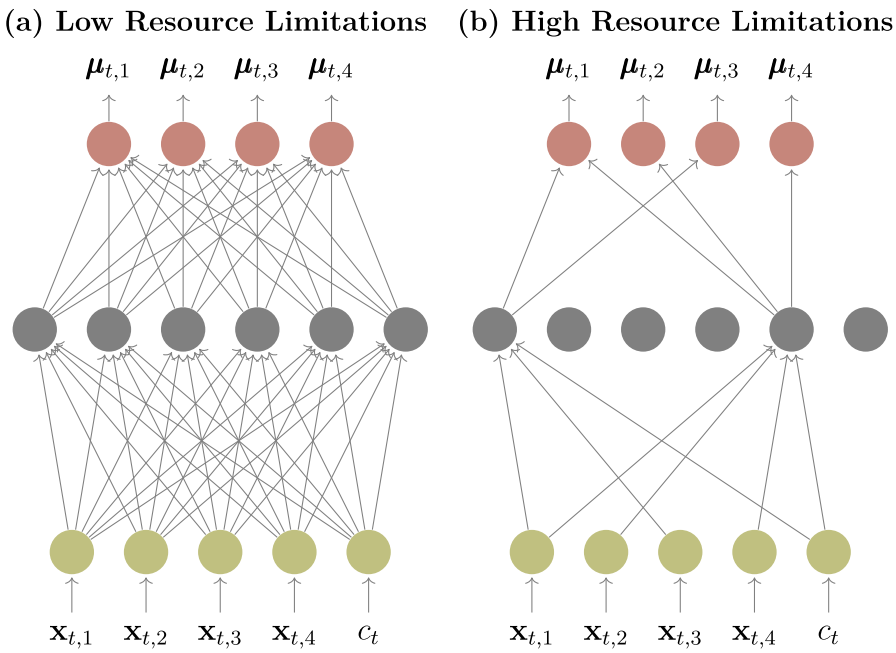
Model Simulations

Next, we demonstrate through a series of model simulations that BMI recovers both single cue and equal weighting heuristics in specific environments. This result implies that both heuristics can be resource-rational strategies under certain conditions. However, we also identify circumstances where BMI does not discover any known heuristic and instead infers strategies that use weighted

[3] The formal justification for this is given by the bits-back argument (Hinton & Van Camp, 1993; Honkela & Valpola, 2004), which allows us to interpret the KL term as the coding length of meta-parameters when encoded together with the data.

Figure 2

Illustration of Two Optimized Neural Networks With a Sparsity-Inducing Prior and Different Resource Limitations



Note. For clarity, we omit recurrent connections and show only the means of $q(\mathbf{w}; \lambda_t)$ as an output. (a) Network trained with low resource limitations uses all available connections. (b) Network trained with high resource limitations uses only the set of connections that are most useful for increasing performance. Network (b) is much simpler than network (a). See the online article for the color version of this figure.

combinations of all features. Before running these simulations, we first have to specify the assumptions we make about the environment and introduce a method for analyzing the emerging strategies.

Environments

For BMI it is necessary to specify a distribution over tasks $p(\mathbf{x}_{1:T}, c_{1:T})$ that is used for meta-learning. In general, this distribution should reflect a participant's prior experiences in the world and its expectations about what tasks might be encountered during the experiment. Here, we make the following assumptions. All tasks involve two options with four different features, and we concentrate on tasks with no costs to reveal information about features. In order to generate a single task, we proceed in three steps:

1. Randomly generate features weights (ref. Equation 1 or 2) by sampling from a standard normal distribution.
2. Randomly generate features $\mathbf{x}_{A,t}$ and $\mathbf{x}_{B,t}$ from a multivariate normal distribution with zero mean and covariance matrix $\sum$.
3. Randomly determine which option has the larger criterion by sampling from a Bernoulli distribution with a success probability given by Equation 2.

Features weights are held constant over a task but are resampled between tasks. Importantly, we assume that the decision-making agent cannot access these weight vectors directly, but instead has to infer them based on observations.

Both redundancy and uncertainty are crucial factors in many real-world decision-making problems (Gigerenzer & Gaissmaier, 2011). Thus, we want them to be present in our environments. Partially redundant features are ensured by drawing separate feature covariance matrices from a prior with $\eta = 2$ (Lewandowski et al., 2009) for each task. To introduce uncertainty, we use a limited number of trials in each task ($T = 10$) and set the additive noise term σ such that an ideal observer is correct in 85% of the cases in the tenth trial.

We consider three variations of the previously outlined environments, that assume (a) known rankings of features, (b) known directions of features or (c) neither. To provide agents with a ranking of features, we arrange them in decreasing order according to the magnitude of their weights. Known directions are ensured by inverting the sign of a feature if it has a negative correlation with the criterion, leading to features with only positive directions.[4] These environments are used during meta-learning, for the model simulation results presented next, and to generate the tasks for our empirical studies.

[4] In our ideal observer implementation, we always assume the original standard normal prior over weights, that is, the prior is not adjusted based on the additional information about ranking or direction. The fact that the prior does not reflect side information makes the resulting model slightly less ideal than a true ideal observer.

Strategy Analysis

To characterize different decision-making strategies, we adopt a measure from the economics literature called the Gini coefficient (Atkinson, 1970). The Gini coefficient was originally intended to describe income and wealth distributions of countries. Its minimal value of zero corresponds to a country in which all residents are equally wealthy, while the maximal value of one corresponds to a country in which a single person possesses everything.[5]

The extreme cases of the Gini coefficient also coincide with the two previously discussed heuristics: Equal weighting heuristics have a Gini coefficient of zero, while single cue heuristics have a Gini coefficient close to one. Thus, we can employ the Gini coefficient to understand how similar estimated regression weights are compared to both heuristics. In practice, we compute Gini coefficients using absolute values of weight vectors. Mathematically, the Gini coefficient of a weight vector $\mathbf{w} \in \mathbb{R}^d$ is defined as half of the relative mean absolute difference, see Equation 9:

$$G(\mathbf{w}) = \frac{\sum_{i=1}^{d}\sum_{j=1}^{d}|\mathbf{w}_i - \mathbf{w}_j|}{2d\sum_{i=1}^{d}\mathbf{w}_i}. \tag{9}$$

Throughout this section, we analyzed Gini coefficients for BMI (with $\beta = 0.01$), MI, and IO. If Gini coefficients were consistently close to zero or one, we deduced that the model has recovered one of the two heuristics. We additionally evaluated the average KL divergence from the posterior predictive distribution of both heuristics to the posterior predictive distribution of BMI. This KL divergence can be interpreted as a difference measure between two models. If it is significantly lower for one of the two heuristics, this would further strengthen our claim that BMI has discovered that particular heuristic.

BMI Discovers Heuristics

First, we considered an environment with known feature rankings. For MI and BMI we optimized meta-parameters until convergence in an environment where features are ordered based on the magnitude for their associated weight. We then analyzed Gini coefficients of inferred regression weights after meta-learning is completed. Because MI and BMI are adapted to the environment, they could exploit the additional ranking information to adjust how they infer strategies.

Figure 3(a) visualizes Gini coefficients obtained from BMI. We observe strategies with nearly maximum Gini coefficients, which correspond to weight vectors that only have a single nonzero component. Thus, we conclude that the single cue heuristic emerged as the resource-rational strategy for an environment with known feature rankings. Looking at MI in Figure 3(b), we find Gini coefficients that cover a much wider range of values. Even though there is an initial tendency toward single cue heuristic, many later decisions are based on compensatory rules. This indicates that being adapted to the environment alone is not a sufficient justification for heuristics. Instead, we need algorithms that are adapted to the environment *and* efficient in terms of their computational resources. Decisions in the ideal observer model are nearly always based on weighted combinations of multiple features, and hence its Gini coefficients in Figure 3(c) spread over an even wider range of values. Figure 3(d) confirms our findings by showing that BMI infers posterior predictive distributions that are much more similar to the single cue heuristic than to equal weighting in terms of their KL divergence.

Next, we looked at an environment where feature directions are known instead of their ranking. For this, we optimized MI and BMI in an environment with only positive feature directions. The result here looks very different compared to the ranking condition. Gini coefficients resulting from BMI, visualized in Figure 4(a), are consistently close to zero. Low Gini coefficients correspond to uniform weight vectors and hence in this environment the equal weighting heuristic turned out to be the resource-rational strategy. Figure 4(b) confirms earlier results showing that MI only leads toward an initial tendency toward heuristics. Early strategies are somewhat similar to equal weighting, but especially as more data is observed strategies with higher Gini coefficients emerge. The ideal observer model on the other hand does not exploit the additional directional information and hence we find no noticeable change in Gini coefficients compared to an environment with known rankings, Figure 4(c). Like before, our results are confirmed when looking at the KL divergence between both heuristics and BMI, which is now substantially smaller for the equal weighting heuristic as shown in Figure 4(d).

BMI Does Not Always Discover Heuristics

We have seen that BMI discovered different heuristics in two classes of environments. Next, we show that there are also environments where this is not the case. For this, we optimized MI and BMI such that they adjust to problems without additional information in the form of ranking or direction. Gini coefficients obtained from BMI reveal that neither single cue nor equal weighting heuristics are resource-rational under such circumstances, as shown in Figure 5(a). Instead, the pattern now looks more similar to one observed in MI and the ideal observer model, shown in Figure 5(b) and (c), respectively. In all cases, Gini coefficients cover the full range of possible values, indicating that inferred weight vectors integrate information from multiple features to different degrees. This time, we find no difference in the KL divergence between both heuristics and BMI, ref. Figure 5(d), which confirms the earlier conclusion that BMI does not recover any of the two heuristics in an environment without additional information about ranking or direction.

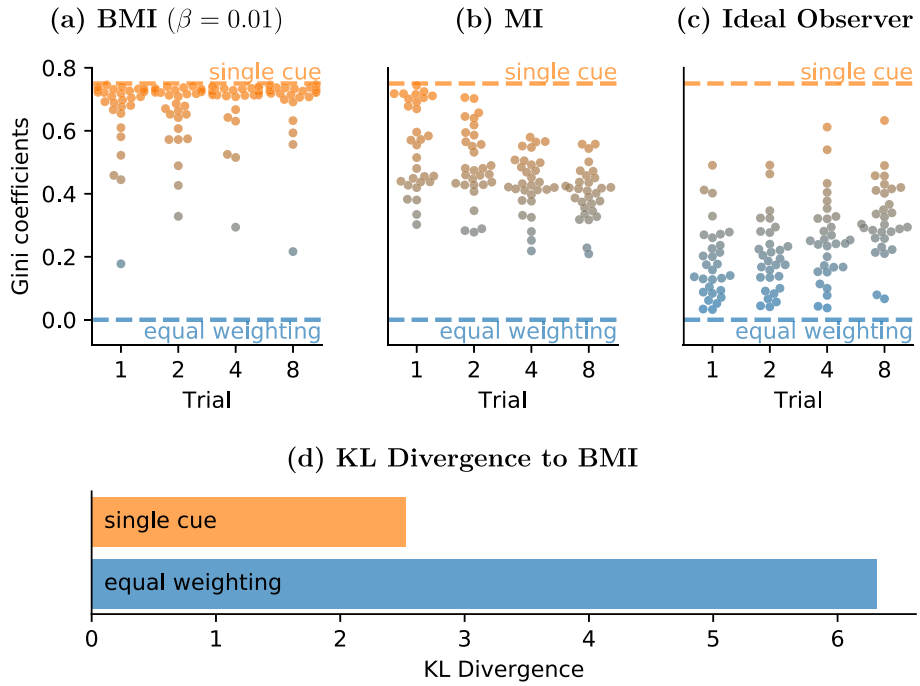
Experimental Predictions

BMI discovered both single cue and equal weighting heuristics when information about ranking and direction was provided, respectively. However, resulting strategies diverged from known heuristics whenever such information was not present. Instead, our simulation results suggest that weighted combinations of multiple features should be used in such situations. Under the assumption that people make adaptive and computationally efficient inferences, our results enable us to make precise predictions about when to expect heuristics as part of human decision-making and when not: Knowing the correct ranking of attributes leads to one reason decision-making, knowing the directions of the attributes leads to equal weighting, and not knowing about either leads to strategies that use

[5] The extreme value of one is only reached in the limit of an infinite number of residents, otherwise the maximum Gini coefficient for d residents is $1 - d^{-1}$.

Figure 3

Strategy Analysis for an Environment With Known Rankings



Note. MI = meta-learned inference; BMI = bounded meta-learned inference; KL = Kullback–Leibler. (a) to (c) Gini coefficients for an environment with known rankings. High values indicate similarity to the single cue heuristic, while low values correspond to equal weighting heuristics. (a) BMI results in Gini coefficients that are close to the single cue heuristic. (b) MI shows tendencies toward the single cue heuristic, especially with few observations. (c) Gini coefficients of the ideal observer model cover the whole range of possible values, indicating that a weighted combination of multiple features is used. (d) Average KL divergence from the posterior predictive distribution of both heuristics to the posterior predictive distribution of BMI. The KL divergence is lower for the single cue heuristic, which confirms our results from the Gini coefficient analysis. See the online article for the color version of this figure.

weighted combinations of multiple attributes. Below, we present the results of three paired comparison studies that confirm the predictions made by BMI.

Experiment 1: Known Ranking

In our first study, participants made decisions in multiple paired comparison tasks while having access to a ranking of features, but not their underlying weights. Previously, we showed that in environments with known feature rankings, single cue heuristics are resource-rational strategies. Hence, we hypothesized that people are more likely to apply the single cue heuristic in this condition.

Method

Participants

Participants were students from the University of Marburg, taking part in the study for course credits. Besides course credits, they got a chance to win a €10 voucher if they made more than 66.6% correct decisions. The experiment was approved by the local ethics board (AZ 2020-32k). In total, we collected data from 28 participants (23 female, average age: 22.36 ± 5.65). We decided on this number of participants based on previous studies (Bröder & Schiffer, 2003; Newell et al., 2003). The median time to complete the experiment was 26.00 min.

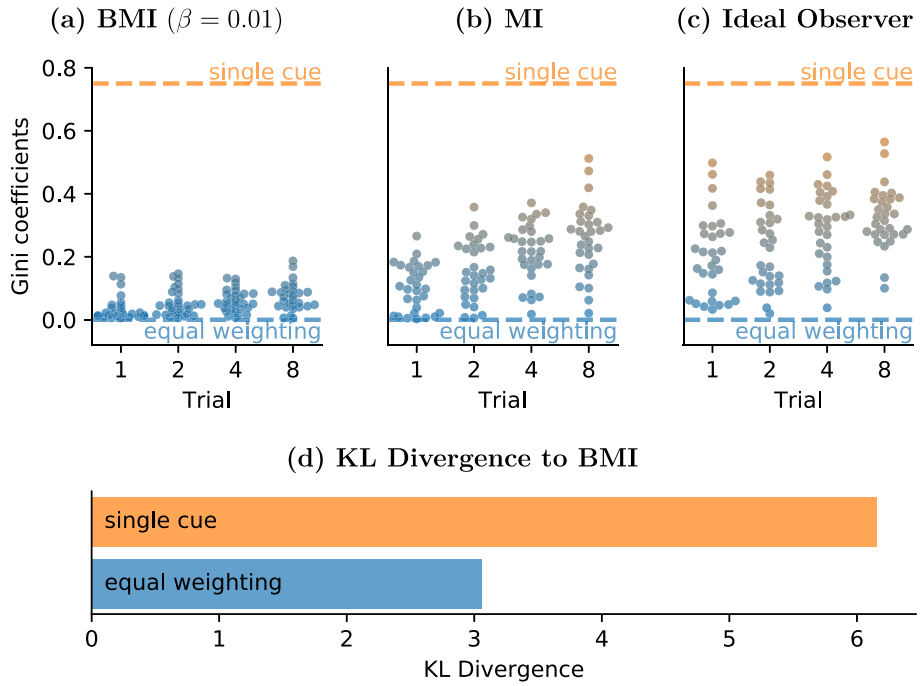
Procedure

Each participant performed 30 different paired comparison tasks that were randomly generated according to the previously described distribution. Each task consisted of 10 trials. The underlying feature weights remained fixed within a task but varied between tasks. Participants were informed about transitions between tasks. Each participant encountered the same set of paired comparison tasks in a randomized order.

The problem itself was framed as an alien sports competition on an unknown planet (see Figure 6). Participants observed four numerical attributes for two aliens and indicated by a button press which alien they believed was more likely to win. The alien cover story was used to keep the meaning of features completely abstract from the participant's perspective. Participants did not have access to the underlying feature weights but instead had to learn about the importance of features based on experience. Feedback about the correct choice was provided directly after each decision. For this condition, features were displayed in descending order based on the magnitude of their weights. Participants were told that features are

Figure 4

Strategy Analysis for an Environment With Positive Directions



Note. MI = meta-learned inference; BMI = bounded meta-learned inference; KL = Kullback–Leibler. (a) to (c) Gini coefficients for an environment with positive directions. High values indicate similarity to the single cue heuristic, while low values correspond to equal weighting heuristics. (a) BMI results in Gini coefficients that are close to the equal weighting. (b) MI shows tendencies toward the equal weighting heuristic, especially with few observations. (c) Gini coefficients of the ideal observer model cover the whole range of possible values, indicating that a weighted combination of multiple features is used. (d) Average KL divergence from the posterior predictive distribution of both heuristics to the posterior predictive distribution of BMI. The KL divergence is lower for the equal weighting heuristic, which confirms our results from the Gini coefficient analysis. See the online article for the color version of this figure.

arranged from top to bottom according to how well they predict the winner. Being aware of this additional ranking information allowed them to apply strategies that are appropriate for this environment. All participants went through a short tutorial and did a comprehension check to confirm that they understood the instructions.

Results

Performance

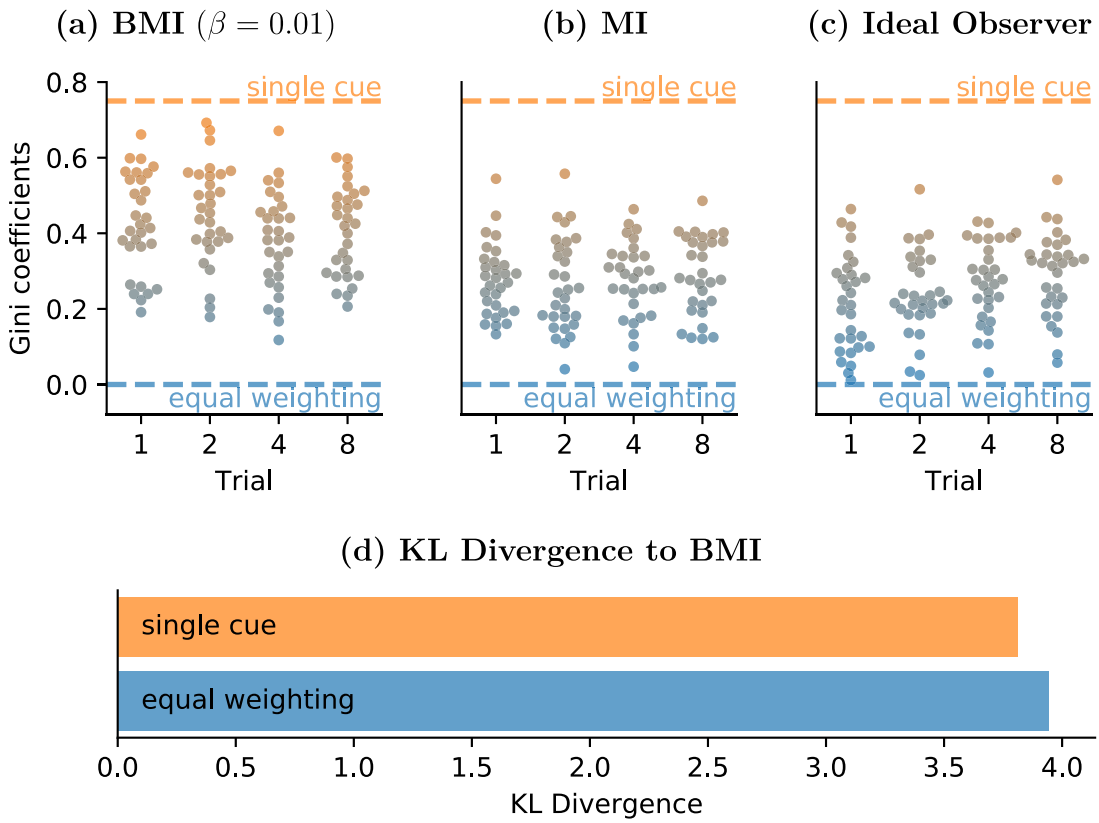
Figure 7 shows the percentage of correct decisions for participants in our study together with the accuracy of different models. Participant performance was within the range of the single cue heuristic but below the ideal observer model. On average, participants made 68.25% ± 7.55% correct choices. For each participant, we assessed whether or not they chose the better option more frequently than chance by using an exact binomial test with a base probability of $p = .5$ and classifying them as better than chance if the p value of that test was smaller than .05. This analysis showed that 26 out of 28 participants chose the better option more frequently than what would be expected under chance level performance. We also fitted a mixed-effects logistic regression to investigate participants' learning over trials and tasks, using a variable indicating whether or not participants had chosen the better option on a given trial as the dependent variable, and adding trial number and task number as both fixed effects and random effects over participants. The results of this model showed a significant fixed effect of trial number ($\beta = 0.12$, $z = 5.63$, $p < .001$) onto choosing the better option but not of task number ($\beta = -0.02$, $z = -0.66$, $p = .51$). This means that participants improved over trials within a given task but did not improve over tasks.

Model Comparison

If people make efficient use of their available computational resources, we expect them to adopt the single cue heuristic in this experiment. To examine this hypothesis, we performed a Bayesian model comparison and computed posterior probabilities of different models given the decisions made by a participant. Appendix D provides a detailed description of the methods we used for statistical analysis. In addition to the previously described models, we also included a simple strategy selection model (ref. Appendix E) and a feedforward network trained by gradient descent (ref. Appendix F) in our analysis. Because the single cue heuristic and BMI make redundant predictions, we decided to split our analysis into two parts. First, we

Figure 5

Strategy Analysis for an Environment Without Ranking or Direction



Note. MI = meta-learned inference; BMI = bounded meta-learned inference; KL = Kullback–Leibler. (a) to (c) Gini coefficients for an environment without ranking or direction. High values indicate similarity to the single cue heuristic, while low values correspond to equal weighting heuristics. (a) BMI, (b) MI, and (c) ideal observer models result in Gini coefficients that cover the whole range of possible values, indicating that a weighted combination of multiple features is used. (d) Average KL divergence from the posterior predictive distribution of both heuristics to the posterior predictive distribution of BMI. The KL divergence is roughly equal for both heuristics, indicating that neither of the two is particularly similar to BMI. See the online article for the color version of this figure.

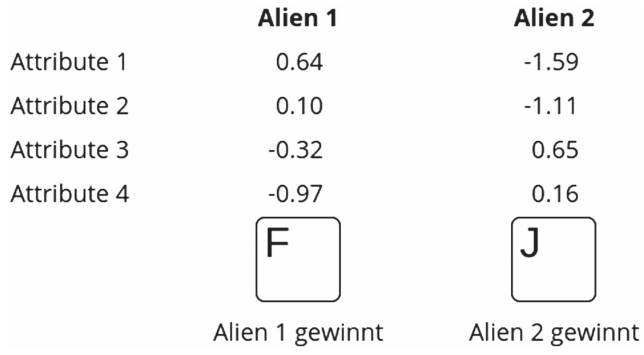
analyzed all models except BMI for individual participants. Then, we compared BMI against the other models on the data of all participants.

We found evidence for the application of the single cue heuristic in 23 out of 28 participants. For all but four of those participants, the model evidence decisively favored the single cue heuristic ($p(m = \text{SC}|\hat{\mathbf{c}}^{(i)}, \mathbf{X}^{(i)}) > 0.99$). Figure 8(a) summarizes posterior probabilities of different models for all participants. From the participants not best described by one reason decision-making, two were best described by guessing, one by the equal weighting heuristic, one by the ideal observer model, and one by the strategy selection model. The protected exceedance probability (PXP), which measures the probability that a particular model is more frequent in the population than all the other models under consideration (Rigoux et al., 2014), favored the single cue heuristic decisively (PXP > 0.999).

While our model simulations predicted that participants should not change their strategy during a task, our analysis did not rule out this possibility so far. It might, for example, be possible that participants did not start a task with the single cue heuristic but only developed this preference during learning. We tested this specific model prediction by comparing differences in log-likelihoods of individual time-steps between different models. Looking at Figure 8(b), we can see that the single cue heuristic dominates both equal weighting and the ideal observer model across all time-steps. This makes it unlikely that participants only applied

Figure 6

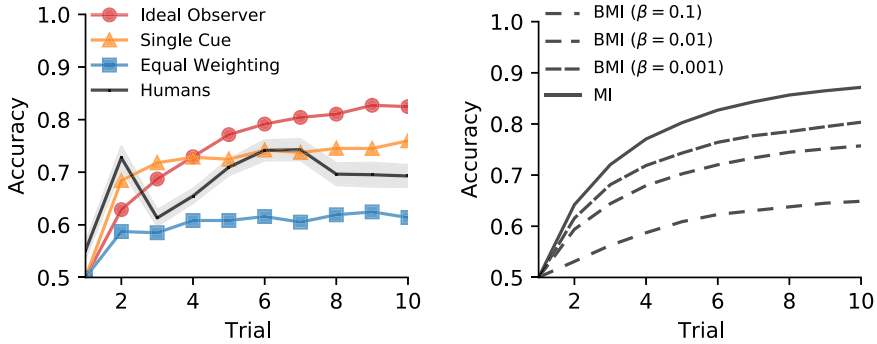
Graphical Illustration of a Single Trial in the Experiment



Note. "Alien *X* gewinnt" translates to "Alien *X* wins"

Figure 7

Percentage of Correct Decisions (Averaged Over All Tasks) in the Ranking Condition Plotted Over the Number of Trials Within a Task



Note. BMI = bounded meta-learned inference. For human performance shaded contours represent the standard error of the mean. The left panel shows the ideal observer model and both heuristics, while the right shows BMI for different values of β. For BMI, lower β-values correspond to a less restricted model. Performance plots for the strategy selection model and feedforward network can be found in Appendices E and F, respectively. See the online article for the color version of this figure.

the single cue heuristic for a subset of trials, and also validates the hypothesis that strategies did not switch within a task.

Finally, we compared how well BMI fared against the other models on the aggregated data. The resulting posterior probabilities indicated that across all participants BMI offered an even better explanation for the observed data than the other models ($p(m = \text{BMI}|\hat{\mathbf{c}}, \mathbf{X}) \approx 1$). We hypothesized that this is the case because BMI was able to explain the behavior of participants that used a single feature (corresponding to higher β-values) *and* those who used two or more features (corresponding to lower β-values). There were overall 12 participants who were better described by BMI than by the single cue heuristic. We found that the fitted β-values of these participants were significantly lower than those within the rest of the population, $t(15.4) = -2.78$, $p = .007$, meaning that these participants acted as if they had access to more resources and therefore applied more complex strategies. There was furthermore a significant rank correlation between β and participants' response times ($\tau = -0.42$, $z = -2.87$, $p = .004$), indicating that people who were better described by models with a shorter description length (i.e., those with higher β-values) also made quicker decisions.

Discussion

Most empirical evidence for one reason decision-making has been provided by studies that involved a cost for acquiring information about features (Bröder, 2000; Bröder & Gaissmaier, 2007; Rieskamp & Otto, 2006). However, even with an experimental protocol that favored few pieces of information, evidence for these strategies remained inconclusive (Newell et al., 2003; Scheibehenne et al., 2013). When information is freely available, people are often better described through compensatory strategies such as logistic regression (Bröder, 2000; Glöckner & Betsch, 2008; Lee & Cummins, 2004; Parpart et al., 2018). Our results are among the first to decisively show that people's choices can be based on a single piece of information, even when such strategies are not favored by the experimental protocol. This was possible because we precisely identified conditions under which one reason decision-making *should* appear. Nearly all participants in our study applied strategies that were efficient in terms of resources while also accounting for environmental characteristics.

Experiment 2: Known Direction

In our second study, we provided no information about ranking and instead informed participants about feature directions; otherwise, it was identical to the first experiment. In our previous analysis, we have seen that this modification also caused a change in what strategy is resource rational. Now, resource-rational decision-making amounts to the application of equal weighting heuristics. We, therefore, hypothesized that participants would become more likely to use such strategies.

Method

Participants

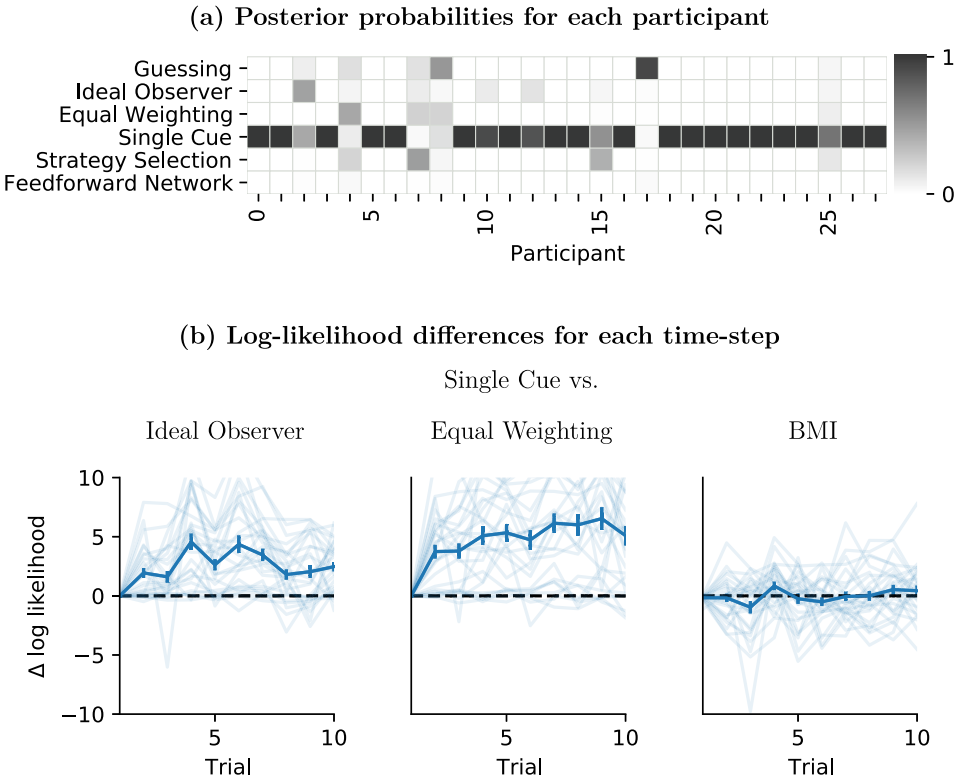
Participants were students from the University of Marburg, taking part in the study for course credits. Besides course credits, they got a chance to win a €10 voucher if they made more than 66.6% correct decisions. The experiment was approved by the local ethics board (AZ 2020-32k). In total, we collected data from 24 participants (22 female, average age: 22.54 ± 3.28). The median time to complete the experiment was 29.69 min.

Procedure

The design was identical to the first experiment, except that participants were informed about the presence of positive feature directions instead of the feature ranking. This was realized by telling them that higher feature values always made it more probable for an alien to win the competition.

Figure 8

Model Comparison in the Ranking Condition



Note. BMI = bounded meta-learned inference. (a) Posterior distributions for each participant over different strategies in the ranking condition. High values indicate that the participant was likely to use the corresponding strategy. (b) Log-likelihood differences for each time-step averaged across all tasks. The solid blue line shows the average across all participants, whereas transparent lines correspond to individual participants. The left panel compares the single cue heuristic to the ideal observer model, the middle panel compares the single cue heuristic to the equal weighting heuristic, the right panel compares the single cue heuristic to BMI. Positive values indicate evidence for the single cue heuristic. See the online article for the color version of this figure.

Results

Performance

Participants made on average 73.85% ± 4.53% correct choices, putting their performance within the range of all models, see Figure 9. The higher average performance indicates that participants found it overall easier to process information about direction than about ranking. We again assessed whether or not individual participants chose the better option more frequently than chance by using an exact binomial test with a base probability of $p = .5$ and classifying them as better than chance if the p value of that test was smaller than .05. This analysis confirmed that all participants chose the better option more frequently than what would be expected under chance level performance. We also fitted a mixed-effects logistic regression to investigate participants' learning over trials and tasks as described in the analysis of the previous study. The results of this model showed a significant fixed effect of trial number ($\beta = 0.08, z = 3.5, p < .001$) onto choosing the better option but not of task number ($\beta = -0.01, z = -0.4, p = .69$). Like in the previous study, this means that participants improved over trials within a given task but did not improve over tasks. Participants' performance in the initial step turned out to be substantially higher than the ideal observer model and both heuristics, indicating that directional information is useful even before making any observations. This characteristic is also captured in BMI.

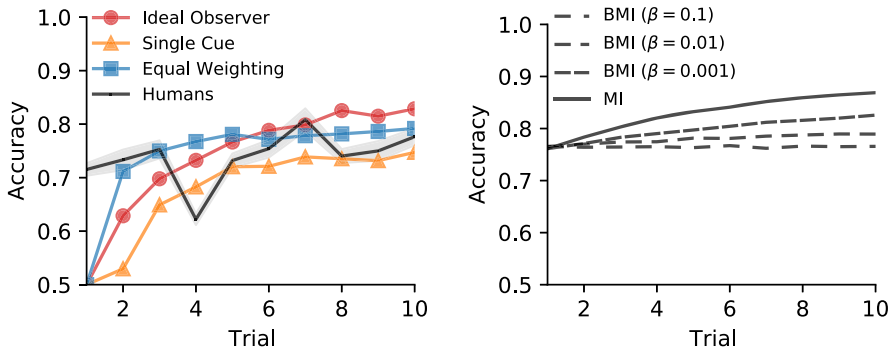
Model Comparison

In this condition, equal weighting and BMI made partially redundant predictions. Thus, we again decided to split our analysis into two parts. First, we analyzed all models except BMI for individual participants. Then, we compared BMI against the other models on the data of all participants.

The posterior probabilities of different models, illustrated in Figure 10(a), confirmed the prediction of our earlier simulations. Participants indeed adhered to the resource-rational maxim and applied equal weighting heuristics. For all participants, equal weighting provided the best explanation for the observed data with decisive evidence ($p(m = \mathrm{EW}|\hat{\mathbf{c}}^{(i)}, \mathbf{X}^{(i)}) > 0.99$). The probability that equal weighting was the most frequent model in the population (PXP > 0.999) supported the conclusion that people, in general, applied equal weighting heuristics when directional information was available. We again inspected per-trial log-likelihoods to confirm that participants did not change their strategies within a

Figure 9

Percentage of Correct Decisions (Averaged Over All Tasks) in the Direction Condition Plotted Over the Number of Trials Within a Task



Note. BMI = bounded meta-learned inference. For human performance shaded contours represent the standard error of the mean. The left panel shows the ideal observer model and both heuristics, while the right shows BMI for different values of β. For BMI, lower β-values correspond to a less restricted model. Performance plots for the strategy selection model and feedforward network can be found in Appendices E and F, respectively. See the online article for the color version of this figure.

task. Figure 10(b) shows that equal weighting dominated the single cue heuristic and the ideal observer model across all time-steps, which again rules out the possibility that participants only applied equal weighting for a subset of trials.

When additionally comparing BMI against the other models on the aggregated data of all participants, we found that BMI again offered an even better explanation than all other models ($p(m = \text{BMI}|\hat{\mathbf{c}}, \mathbf{X}) \approx 1$). This was the case because BMI was able to capture participants' decisions in the initial step, while the equal weighting heuristic did not. This can be confirmed by inspecting the rightmost panel of Figure 10(b), which compares per-trial log-likelihoods between equal weighting and BMI. Here, we find a substantial difference between how well both strategies matched human choices in the initial trial, but no differences during later trials. Indeed, if we look at how well BMI describes participants on an individual level, we find that it offers a better explanation than equal weighting for all but four participants. We again found a significant rank correlation between β and participants' response times ($\tau = -0.36$, $z = -2.16$, $p = .03$), confirming the result from our first study showing that people who were better described by algorithms with a shorter description length (i.e., those with higher β-values) also made quicker decisions.

Discussion

Like in our first study, we found that people apply resource-rational strategies that are adequate for the given environment. Participants performed better compared to the first study, indicating that they found it easier to work with directions than with rankings. We speculate that one explanation for this observation could be that positive correlations are more frequently encountered in the world. Perhaps somewhat surprisingly, there is only limited evidence from prior decision-making studies showing that people employ equal weighting heuristics. The present study is amongst the first to show that people rely heavily on such strategies under the appropriate conditions. However, there is a result from the MCPL literature that connects nicely to our result. In this context, Newell et al. (2009) showed that people also switched to an equal weighting heuristic when provided with directional information about feature weights.

Experiment 3: Unknown Ranking and Direction

In our final study, we investigated choice behavior in an environment that did not provide information about ranking or direction. In the previous model simulations, we have demonstrated that no heuristic emerges under such conditions. Instead, BMI discovered strategies with compensatory weights even under large resource constraints. Hence, we predicted that people in this condition should be less reliant on traditional heuristics and instead integrate information from multiple features properly. To test this hypothesis, we initially reiterated the experimental paradigm of our two earlier studies. However, we found that without the additional information about ranking or direction, cognitive limitations became a dominating factor. While some participants still performed well, a substantial number were at or close to chance level. Therefore, we subsequently decided to conduct a simpler version of our task which involved only two features. While we focus on the two-feature study in the main text, the results of the four-feature study are also reported in Appendix G for completeness.

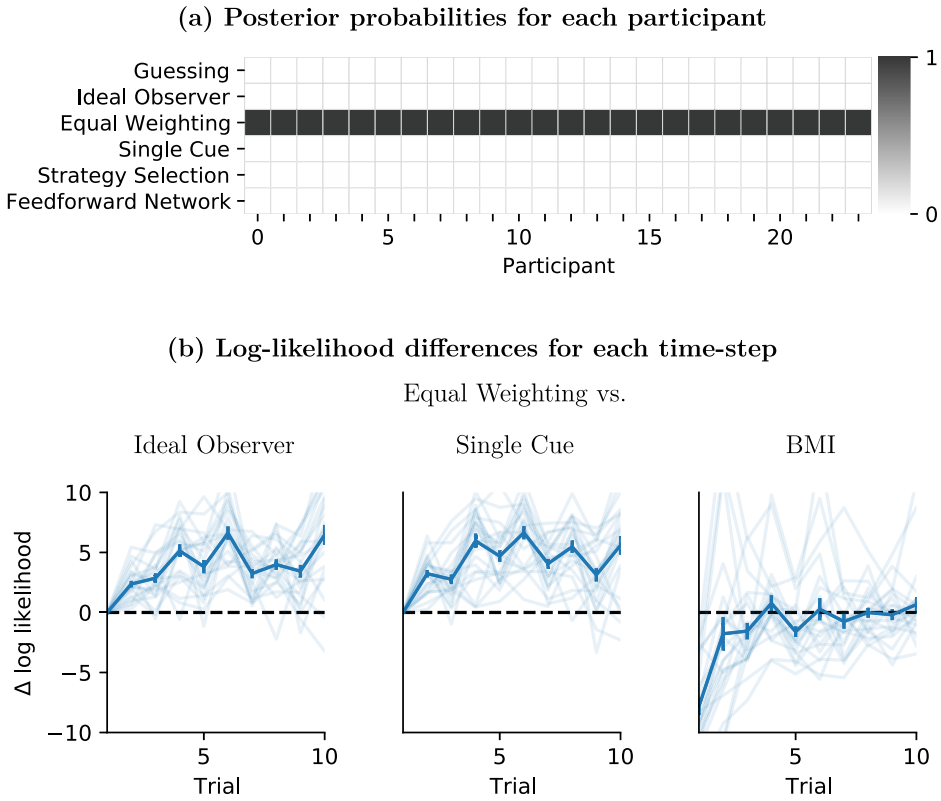
Method

Participants

Participants were students from the University of Marburg, taking part in the study for course credits. Besides course credits, they got a chance to win a €10 voucher if they made more than 66.6% correct decisions. The experiment was approved by the local ethics board (AZ 2020-32k). In total, we collected data from 27 participants (22 female, average age: 21.74 ± 4.75). The median time to complete the experiment was 36.35 min.

Figure 10

Model Comparison in the Direction Condition



Note. BMI = bounded meta-learned inference. (a) Posterior distributions for each participant over different strategies in the direction condition. High values indicate that the participant was likely to use the corresponding strategy. (b) Log-likelihood differences for each time-step averaged across all tasks. The solid blue line shows the average across all participants, whereas transparent lines correspond to individual participants. The left panel compares equal weighting to the ideal observer model, the middle panel compares equal weighting to the single cue heuristic, the right panel compares equal weighting to BMI. Positive values indicate evidence for the equal weighting heuristic. See the online article for the color version of this figure.

Procedure

The general design was identical to both previous experiments, except for two adjustments: Participants only observed two features for each alien, and they were not provided with information about the feature ranking and their directions anymore. We additionally probed participants' judgments about ranking and direction of features at the end of each task. In particular, we asked them for both attributes whether they believe a positive value is advantageous for winning the competition, and which of the two attributes they believe is more important for determining the winner. For all questions, we collected a binary response.

Results

Performance

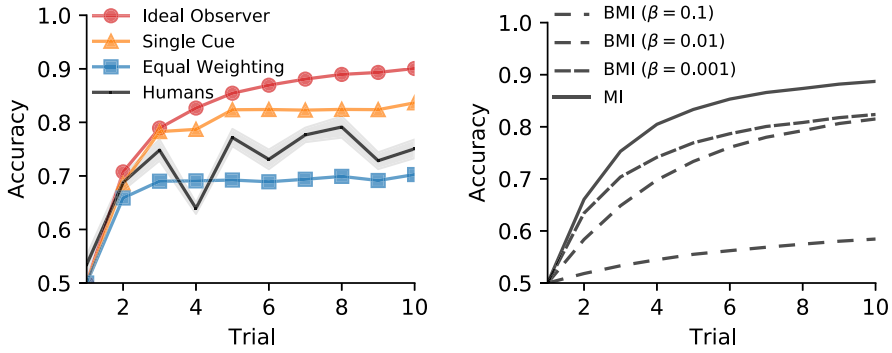
The average performance of participants was 71.59% ± 5.68%, which places them somewhere between the single cue and equal weighting heuristics (see Figure 11). Like in the two previous studies, we assessed whether or not individual participants chose the better option more frequently than chance by using an exact binomial test with a base probability of $p = .5$ and classifying them as better than chance if the p value of that test was smaller than .05. This analysis indicated that 26 of 27 participants chose the better option more frequently than what would be expected under chance level performance. We also repeated the mixed-effects logistic regression used in the previous two studies to investigate participants' learning over trials and tasks. This analysis revealed a significant fixed effect of trial number ($\beta = 0.2$, $z = 8.83$, $p < .001$) onto choosing the better option but not of task number ($\beta = 0.05$, $z = 1.3$, $p = .19$), again meaning that participants improved over trials within a given task but did not improve over tasks.

Model Comparison

According to our model simulations, we should expect to find evidence for models using weighted combinations of multiple features in this condition. Because no known heuristic emerged in this environment, we did not split our analysis and already

Figure 11

Percentage of Correct Decisions (Averaged Over All Tasks) in the Unrestricted Condition Plotted Over the Number of Trials Within a Task



Note. BMI = bounded meta-learned inference. For human performance shaded contours represent the standard error of the mean. The left panel shows the ideal observer model and both heuristics, while the right shows BMI for different values of β. For BMI, lower β-values correspond to a less restricted model. Performance plots for the strategy selection model and feedforward network can be found in Appendices E and F, respectively. See the online article for the color version of this figure.

considered BMI on the level of individual participants. Posterior probabilities obtained from a Bayesian model comparison in Figure 12(a) confirmed that most participants combined information from multiple features instead of using heuristics like equal weighting or one reason decision-making. Twenty two out of 27 participants were best described by BMI; in 16 of those we found decisive evidence ($p(m = \text{BMI}|\hat{\mathbf{c}}^{(i)}, \mathbf{X}^{(i)}) > 0.99$). Amongst the participants not best described by BMI, three were best described by the ideal observer model and two by the equal weighting heuristic. We again found that BMI fared favorably against all other models on the aggregated data ($p(m = \text{BMI}|\hat{\mathbf{c}}, \mathbf{X}) \approx 1$). The protected exceedance probability (PXP > 0.999) also supported the conclusion that BMI was the most frequent explanation for participants in our population. Looking at per-trial log-likelihoods in Figure 12(b), we see that BMI dominated all alternative hypotheses on nearly every time-step, which again confirms that participants did not change their strategies within a task. Like in the first two studies, there was a significant rank correlation between β and participants' response times ($\tau = 0.37$, $z = -2.54$, $p = .01$), confirming that people who were better described by algorithms with a shorter description length (i.e., those with higher β-values) generally made quicker decisions.

Judgments About Ranking and Direction

What made BMI a better model of human choices than other compensatory strategies like the ideal observer model? We speculated that this was the case because BMI better reflected human intuitions about the ranking and direction of features. To test this hypothesis, we analyzed participants' judgments about ranking and direction at the end of each task. For our analysis, we computed likelihood ratios of human judgments between the two competing models. The likelihood that model $m \in \{\text{BMI}, \text{IO}\}$ assigns a positive direction to feature i is given by $p(\mathbf{w}_i > 0|\mathbf{x}_{1:T}, c_{1:T}, m)$, which can be computed in closed form under our assumption of normal posterior distributions. The likelihood that model $m \in \{\text{BMI}, \text{IO}\}$ evaluates the first feature as more important is given by $p(|\mathbf{w}_1| > |\mathbf{w}_2||\mathbf{x}_{1:T}, c_{1:T}, m)$, which we approximated using a sample-based estimate. While we found no substantial difference in terms of ranking (BF = 0.41), BMI offered a much better explanation for the human judgments of directions (BF = 1.4×10^{24}). This suggests that the main advantage of BMI stems from the fact that it is better at capturing human intuitions about feature directions.

Discussion

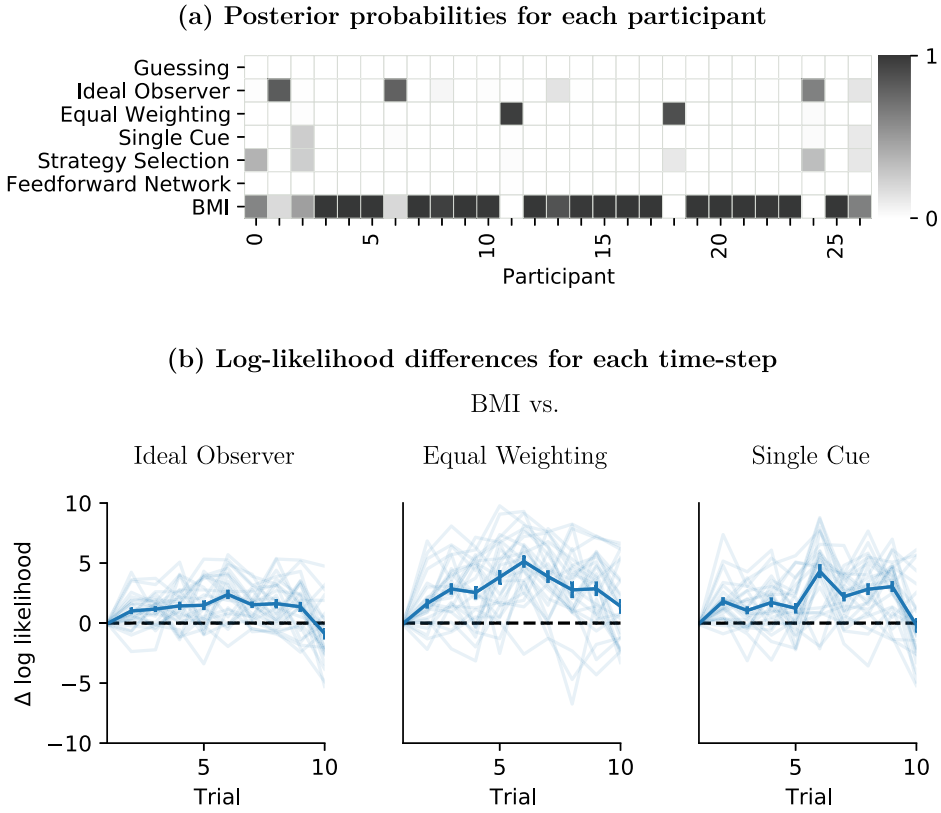
In an environment that did not provide additional information about ranking or direction, participants' decision-making again followed the prediction made by BMI. The majority of participants applied strategies that involved weighted combinations of features, as was suggested by our model simulations. We observed an identical pattern in our initial study with four features, but there it was less pronounced due to the complexity of the task (see Appendix G). The general result that most people were able to quickly combine information from multiple sources if needed is also consistent with results of prior studies (Bröder, 2000; Glöckner & Betsch, 2008; Parpart et al., 2018). Notably, BMI offered a superior account to alternative compensatory strategies like the ideal observer model. We believe that part of the explanation for this result is that BMI aligned better with the subjective judgments about feature directions than other compensatory strategies. However, there might be additional factors at play, which our current form of analysis was not able to capture.

General Discussion

At the core of theories of ecological rationality, researchers have posited an interaction between cognition and the environment. Brunswik (1956) argued that human perception cannot be understood in laboratory settings alone, but rather has to be interpreted in the light of real environments in which real objects are perceived and acted upon. Simon (1990b) famously highlighted the interaction between cognition and the environment using an analogy of a pair of

Figure 12

Model Comparison in the Unrestricted Condition



Note. BMI = bounded meta-learned inference. (a) Posterior distributions for each participant over different strategies in the unrestricted condition. High values indicate that the participant was likely to use the corresponding strategy. (b) Log-likelihood differences for each time-step averaged across all tasks. The solid blue line shows the average across all participants, whereas transparent lines correspond to individual participants. The left panel compares BMI to the ideal observer model, the middle panel compares BMI to equal weighting, the right panel compares BMI to the single cue heuristic. Positive values indicate evidence for BMI. See the online article for the color version of this figure.

scissors, with one blade being the structure of the environment and the other blade the computational capabilities of the subject. This conceptualization of ecological rationality has strongly influenced theories of heuristic decision-making. The need to economize cognitive resources places pressure on the mind to employ heuristics that work well in specific environments. Nonetheless, how people pick a particular heuristic for a specific environment and where those heuristics come from in the first place has remained elusive. The theoretical picture becomes even more puzzling when looking at the empirical support for heuristic decision-making. Proponents of heuristic decision-making acknowledge these problems. For example, Gigerenzer (2008) writes: Why do heuristics work? They exploit evolved capacities that come for free. In addition, they are tools that have been customized to solve diverse problems. By understanding the ecological rationality of a heuristic, we can predict when it fails and succeeds. The systematic study of the environments in which heuristics work is a fascinating topic and is still in its infancy. But what does a theory, which can explain how heuristics emerge and how they are selected, look like? We have put forward BMI as a theory that makes significant advances on these questions. Our simulation results show that BMI discovers previously suggested heuristics. Thus, it provides a normative justification for heuristic decision-making. Moreover, we find that different heuristics emerge depending on environmental assumptions. Thus, BMI also explains how decision-making strategies are selected.

Already early on, researchers working on heuristic decision-making levied the criticism that simply observing behavioral biases is not enough, and that in place of plausible heuristics that explain everything and nothing—not even the conditions that trigger one heuristic rather than another—we need models that make surprising (and falsifiable) predictions (Gigerenzer, 1996). However, the very fact that several heuristic components have been claimed to be part of a heuristic toolbox without fully specifying how they are selected and combined, has subjected heuristic theories to a similar line of criticism: "... if one cannot predict which heuristics will be used in which environments then determining the heuristic that will be selected from the toolbox for a particular environment becomes necessarily post hoc and thus the fast-and-frugal approach looks dangerously like becoming unfalsifiable." (Newell et al., 2003). In contrast to these arguments, BMI makes clear, falsifiable, and surprising predictions about when people should apply which

heuristic. Specifically, our simulation results show that there are three important classes of environments triggering three decision-making strategies. If people know the correct ranking of attributes but not their weights, then they should exhibit one reason decision-making. If people know the direction of the attributes but not their ranking, then they should exhibit equal weighting strategies. Finally, if people do not know either the ranking or the direction of the attributes, then they should exhibit strategies that use weighted combinations of attributes.

We subjected these predictions to a rigorous test in three paired comparison experiments and found that the vast majority of participants applied decision-making strategies as predicted by BMI. Moreover, BMI captured elements of human decision-making that could not be explained by traditional heuristics in all three experiments: In the first study, it additionally accounted for the participants that employed more complex strategies. In the second study, it provided an explanation for the good initial performance of participants. In the third study, it predicted correctly that people make decision using a weighted combination of all features, and offered a superior account to alternative compensatory strategies like the ideal observer model. These results enrich our theoretical and empirical understanding of ecologically rational decision-making.

Limitations

Gigerenzer and Todd (1999) argue that decision-making under limited resources cannot be expressed through models that perform optimization under constraints: Optimization under constraints also limits search, but does so by computing the optimal stopping point, that is, when the costs of further search exceed the benefits. Computing this optimal stopping point can be at least as expensive as finding the optimal solution; hence it defeats the initial intention of modeling decision-making under resource limitations (Gigerenzer & Todd, 1999; Scheibehenne & von Helversen, 2009). BMI involves optimization under constraints but importantly does so at the meta-learning level, which happens on a much larger time scale (e.g., through evolutionary processes). Learning within an individual task, on the other hand, is fast as it does not involve any form of optimization. This perspective of learning at multiple scales is also at the core of recent theories of fast and slow reinforcement learning (Botvinick et al., 2019).

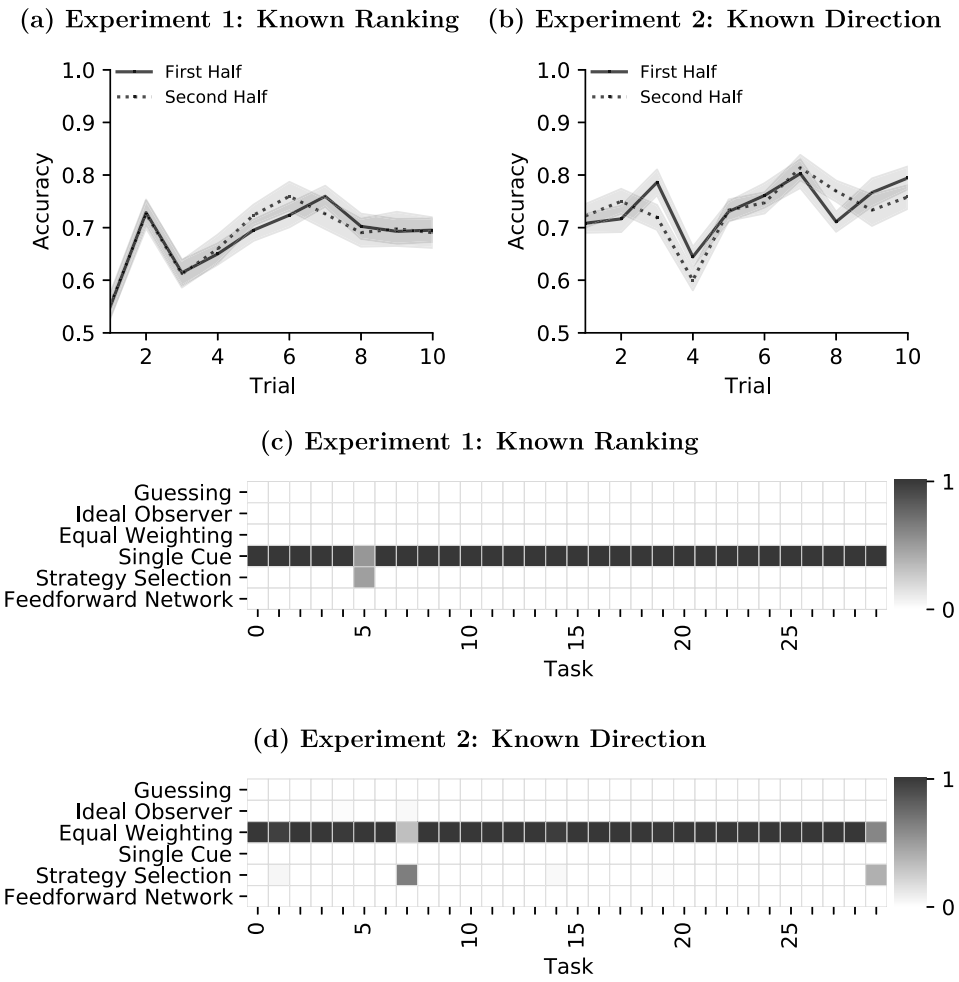
BMI assumes that meta-learning happened prior to the experiment, but it remains agnostic about the exact processes controlling the acquisition of strategies. BMI could, for example, be acquired through evolutionary processes, individual experiences, or both. If meta-learning indeed happened prior to the experiment, we should find no noticeable improvement in performance throughout our studies. We find support for this hypothesis when comparing human performance in the first and second half of our studies as shown in Figure 13(a) and (b). Furthermore, we evaluated posterior probabilities of different models for each task as opposed to for each participant, shown in Figure 13(c), and found that participants did not switch between different strategies during the experiment. Nonetheless, a valid criticism of our current work is that it does not address the precise process of meta-learning and whether this process is rather shaped by ontogeny, phylogeny, or both. This is indeed an open problem for all theories of heuristic decision-making, which at various times have argued that heuristics emerge from evolutionary pressures (Hutchinson & Gigerenzer, 2005), developmental processes (Gigerenzer, 2003), or task-specific adaptations (Marewski & Schooler, 2011). The time scale of meta-learning, therefore, remains an open theoretical and empirical question.

We have used a particular model architecture to simulate behavior in our tasks. In particular, we applied a gated recurrent network and adapted the meta-parameters through gradient descent on a loss function that can trade-off between the accuracy of the network and the description length of its parameters. Thus, a naturally arising question is how much our results depend on the chosen architecture. For the sake of the resource-rational argument, we should have used the architecture that optimally solves the accuracy-effort trade-off. Because identifying this architecture is not possible, we settled for the next best option and used an architecture that is known to work well across a wide range of domains. Theoretically, a resource-rational algorithm should at least be able to recover optimal decision-making if there are no resource limitations. Infinitely wide recurrent neural networks are known to be Turing-complete and hence are in theory able to implement optimal decision-making (Siegelmann & Sontag, 1992). Looking at Figure 11, we observe that our networks are wide enough to closely approximate the ideal observer model.

We have also used a particular distribution over tasks for training our meta-learning models. Even though we constructed this distribution to reflect real-world decision-making environments, it remains unclear whether our assumptions can be fully justified. This is a general criticism that rational accounts of decision-making must face (Binmore, 2007; Brighton & Gigerenzer, 2012; Gigerenzer & Gaissmaier, 2011; Szollosi & Newell, 2020). For example, Szollosi and Newell (2020) argued that a fundamental flaw in theories that rely on the intuitive statistician metaphor [including rational accounts] is that the correspondence problem [i.e., how to develop a representation of the environment] is solved in advance by the researcher. However, BMI does not necessarily require a researcher to specify what the environment should look like. In principle, it only needs examples of tasks from the environment. Thus, BMI can sidestep the above criticism by using a meta-learning distribution that is constructed from *actual* real-world decision-making tasks. How to construct such a meta-learning distribution is an interesting question, which we hope to address in future work. Furthermore, we only investigated learning over a rather small number of trials per task. Therefore, we might not have detected all possible learning effects that could occur with a more prolonged learning phase. For example, we do not believe people are, in general, unable to perform well in the more complex, four-feature version of our third study and that increasing the number of trials per task could facilitate learning to the extent that people perform adequately at the given task. Finally, it would be interesting to investigate whether the insights gained from our studies can be transferred to different task formats, for example, causal learning (Lagnado et al., 2013; Waldmann & Holyoak, 1992) or active learning (Gureckis & Markant, 2012; Parpart et al., 2017).

Currently, our approach also does not directly offer a way to predict which properties of the environment will determine what type of decision-making strategies are ecologically rational. Instead, we have to train our meta-learning models in different environments and then analyze what decision strategies emerge, for example by analyzing the weights' Gini coefficient. Looking at a model's emerging properties is a common method when neural network approaches are applied to psychological questions (Ritter et al., 2017). We believe that this possible weakness can also be a strength, because it forces researchers to truly study the properties of

Figure 13
Behavioral Development During the Experiment



Note. (a) and (b) show that the performance of participants did not change over the experiment, indicating that meta-learning already happened prior to the experiment. Shaded contours represent the standard error. (c) and (d) confirm this observation by showing that the selection of strategies also did not change during the experiment. High values indicate that the corresponding strategy was applied with high probability in the given task.

environments, as has been the core proposal of theories of ecological rationality for decades.

Related Work

To highlight what BMI adds to existing theories, we compare it to other ideas put forward in previous investigations. In the context of decision-making, we focus on methods that address how strategies are selected and how they are discovered. Beyond that, we discuss how meta-learning and resource rationality have been applied to understand other phenomena.

Strategy Discovery

There have been some accounts that explain how strategies are discovered. Schulz et al. (2016) proposed a method for learning decision-making strategies from small, probabilistic building blocks. Based on a self-reinforcing sampling scheme, they were able to build tree-like noncompensatory heuristics. Their approach can recover TTB on data-sets that have been generated by the TTB heuristic. However, it is not able to learn about other, noncompensatory strategies or to make predictions about when participants would prefer which strategy.

Lieder et al. (2017) suggested a model that composes strategies from atomic computations. According to their theory, an agent represents computations as costly actions in a metalevel Markov decision process. The agent's goal is to maximize the external payoff obtained from making correct decisions while accounting for the computational costs of actions. When they applied their theory to several decision-making problems, they found that it discovered two known heuristics—TTB and guessing—as well as a novel strategy that combined TTB with satisficing (Simon, 1956).

Parpart et al. (2018) showed that heuristics can emerge from Bayesian inference in the limit of infinitely strong priors. Using this

idea, they identified priors corresponding to an equal weighting heuristic. Finding a prior that leads to TTB proved to be more challenging in the Bayesian framework and was only possible after introducing an additional decision rule. Instead of relying on the complexity argument as justification for heuristics, their analysis suggested that heuristics work well because they implement priors that reflect the actual structure of the environment.

Theories that build algorithms from simpler computations (Lieder et al., 2017; Schulz et al., 2016) discover one reason decision-making heuristics without difficulties, but struggle to account for equal weighting heuristics. Theories based on Bayesian inference (Parpart et al., 2018) on the other hand have no difficulties with discovering equal weighting heuristics, but require additional components to find heuristics that rely on a single piece of information. We have shown that people use both classes of strategies and provided a theory that can discover both of them in an appropriate context.

Strategy Selection

There have also been several theories explaining how strategies are selected. Rieskamp and Otto (2006) proposed a theory of strategy selection learning that framed the strategy selection process as a model-free reinforcement learning problem. Their theory assumes that people slowly learn how to select the right strategy from a given repertoire of strategies based on repeated interactions. A key finding of their experiments was that over time participants learned to select the best-performing strategy for a particular environment. Their method requires learning from scratch whenever it encounters novel problems and hence it does not address how knowledge is transferred between different environments, and why participants are immediately able to select appropriate strategies in our experiments.

Lieder and Griffiths (2017) addressed the missing ability to transfer knowledge between environments through an approach based on rational metareasoning. Based on properties of the environment, they predicted speed and accuracy of different strategies. They showed that participants selected the strategy that was best for solving the speed-accuracy trade-off in the current context. In contrast to their work, we used separate models for each environment. However, it would be possible to extend our modeling framework by conditioning the initial state of the recurrent network on features of an environment.

Marewski and Schooler (2011) postulated a probability landscape describing an individual's ability to apply a strategy as a function of cognitive capabilities and the environment. Their work referred to situations in which a strategy can be applied as a cognitive niche and showed that cognitive niches of different strategies are disjoint in many cases. This greatly simplified the strategy selection problem and was in line with participants' behavior across a number of experiments. We believe that cognitive niches could also be the result of meta-learning, where an algorithm adapts to a given characteristic of an environment until it cannot easily be applied to a vastly different environment anymore.

Previous theories of strategy selection require defining a set of potential strategies in advance, which can be problematic because it always comes along with the risk of missing out on the strategy that is appropriate for solving the problem at hand. In contrast, BMI is not restricted to predefined sets and instead discovers useful strategies on the fly. While there exist prior approaches that address either the strategy selection problem or the strategy discovery problem independently, BMI is also the first to account for both problems jointly within a unified framework.

Resource Rationality

The space of existing resource-rational models is large and such models have been applied to study human behavior across a wide range of contexts (for extended summaries on this topic see Bhui et al., 2021; Gershman et al., 2015; Lieder & Griffiths, 2019). In this subsection, we provide a brief review of such models to highlight how our approach relates to the previous literature. There exist many conceptualizations of what constitutes a computational resource. Two of the most common ones are *computation time*, that is, the number of steps necessary to solve a problem, and *storage space*, that is, the amount of memory required for solving a problem.

Lieder et al. (Lieder & Griffiths, 2017; Lieder et al., 2017, 2018) proposed to model limited computation time as a form of rational metareasoning (Russell & Wefald, 1991). They defined the value of computation as the difference between the utility of a strategy and its execution time and argued that a resource-rational agent should maximize this quantity. This approach is extremely general and can also be applied to costs other than computation time. However, it requires designing an appropriate cost function for the specific problem at hand. Another way to restrict computation time is offered by sampling-based models (Ortega et al., 2015; Sanborn et al., 2010; Vul et al., 2014). In such models, ideal inference is approximated through Monte Carlo sampling, and decreasing the number of samples is interpreted as a reduction in computation time.

Limited storage space, on the other hand, is typically modeled through methods that appeal to rate-distortion theory (Bates & Jacobs, 2020; Genewein et al., 2015; Gershman, 2020; Ho et al., 2020; Sims, 2018; Zaslavsky et al., 2018). In this framework, one attempts to maximize some measure of performance, while simultaneously placing an upper bound on the number of bits required to store an object of interest. What kind of performance measure is maximized and what kind of object is stored depends on the specific model instantiation. Zenon et al. (2019) identified two major classes of cognitive costs that can be represented using rate-distortion theory: (a) a perceptual cost for storing representations of stimuli, and (b) a cost for storing deviations from the default behavior. In our setting, these two costs translate to a cost for storing a representation of feature vectors and a cost for storing deviations from the default decision-making strategy, respectively.

BMI is similar to traditional rate-distortion theory-based approaches as it also places a cost on storage space. However, it neither implements a cost for storing feature vectors, nor a cost for storing deviations from the default strategy. Instead, it places a storage cost on the algorithm that infers which decision-making strategies to apply. In some sense, this is similar to the concept of Kolmogorov complexity (Chaitin, 1969; Kolmogorov, 1965; Solomonoff, 1964), which measures the size of the shortest computer program that produces an object of interest. Kolmogorov complexity has previously been applied to the study of cognition (Chater & Vitányi, 2003; Gauvrit et al., 2014, 2017; Griffiths et al., 2018; Zenil et al., 2015). However, because Kolmogorov complexity is based on universal programming languages, it comes with the downside of being incomputable in general. BMI relaxes the assumption of universal programming languages, and could therefore be viewed as a practical implementation of Kolmogorov complexity.

Meta-Learning in the Context of Human Behavior

Brighton (2006) and Chater et al. (2003) considered standard feed-forward networks trained with backpropagation as models of decision-making in paired comparison tasks. Their results indicated that, if only a few examples were used, such models tended to overfit and were outperformed by much simpler, more robust alternatives. Brighton (2006) suggested meta-learning as a potential solution to this problem of overfitting but did not provide a concrete implementation of this conjecture. BMI is such an implementation that can be applied to paired comparison tasks with few examples and—crucially—*without* showing signs of overfitting. The key to BMI's success is that learning happens solely in the fully-trained network's recurrent activations and not through traditional gradient-based training schemes.

When we look beyond decision-making and paired comparison tasks, meta-learning has recently received increased attention as an explanation for human behavior across a variety of cognitive and neuroscientific questions. For example, meta-learning has been shown to lead to human-like characteristics in the contexts of few-shot learning (Santoro et al., 2016), systematic compositionality (Lake, 2019), exploration (Binz & Endres, 2019) as well as one-shot navigation and model-based reasoning (Wang et al., 2016). Most relevant to our work is the approach of Dasgupta et al. (2020), who taught neural networks to approximate Bayesian inference, given some information about an inference problem's prior and likelihood. They are able to account for a large number of cognitive biases, including base rate neglect and conservatism, by restricting the size of the network. This approach shares its core principles with our theory: resource rationality and meta-learning. However, BMI does not approximate Bayesian inference explicitly as done by Dasgupta et al. (2020). Instead, it attempts to infer distributions that are optimal for making predictions. In the limit of no resource limitations, this also leads to algorithms that approximate Bayesian inference (Ortega et al., 2019). However, when computational resources are limited, the two approaches will produce algorithms with distinctive characteristics.

Future Directions

Most computational models in psychology and cognitive science are confined to idealized settings. BMI on the other hand can—in principle—scale to much more complex domains (Santoro et al., 2016; Wang et al., 2016). Having access to such models allows us to study human behavior under more realistic conditions. In the context of decision-making, it becomes, for example, possible to investigate how and why different representational formats influence human strategies (Bröder & Schiffer, 2006) by learning models that directly process visual representations of the task.

The classical approach to computational modeling is to propose a model, test its predictions and finally revise the model if required. However, we can also envision an approach for the revision of theories that puts the study of environments first. In this framework, we would ask ourselves what environments can account for observed behavior assuming that people make ecologically and resource-rational decisions, instead of revising arbitrary parts of the model. That this is a promising research direction for building more human-like agents was shown for example by Hill et al. (2020), who demonstrated that systematic generalization can be an emergent property of an agent interacting with a *rich* environment.

Finally, our theory provides us with a set of predictions about what should happen when available computational resources are manipulated. It will be interesting to see whether people follow the behavioral trajectories stipulated by BMI when put under cognitive load or whether patients with attention or memory impairment are better described by models with lower complexity.

Conclusion

The idea that theories of human cognition should consider both the structure of the environment and the computational capabilities of the subject has been a central theme in psychology (Simon, 1990b; Todd & Gigerenzer, 2012). However, actual implementations of this principle have been lacking so far. BMI provides such an implementation by combining the ideas of resource rationality and meta-learning. BMI accounts for two open questions in the decision-making literature simultaneously, explaining why different strategies emerge and how appropriate strategies are selected. By mapping out environments that cause different strategies to be resource-rational, we obtained precise predictions about when previously suggested heuristics should be used and when not. We confirmed these predictions in three paired comparison experiments. Taken together, we believe that BMI offers a normative and empirically supported theory of human decision-making.

References

Achille, A., & Soatto, S. (2018). Emergence of invariance and disentanglement in deep representations. *The Journal of Machine Learning Research*, *19*(1), 1947–1980.

Achille, A., & Soatto, S. (2019). *Where is the information in a deep neural network?* arXiv preprint arXiv:1905.12213.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning.*Annual Review of Psychology*, *56*, 149–178.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263.

Ayal, S., & Hochman, G. (2009). Ignorance or integration: The cognitive processes underlying choice behavior. *Journal of Behavioral Decision Making*, *22*(4), 455–474.

Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological Review*, *127*(5), 891–917.

Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multiattribute choice. *Operations Research*, *56*(5), 1289–1304.

Bengio, Y., Bengio, S., & Cloutier, J. (1991). Learning a synaptic learning rule. *IJCNN-91-Seattle International Joint Conference on Neural Networks* (Vol. *2*, pp. 969).

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 107–129.

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

Binmore, K. (2007). Rational decisions in large worlds. *Annals of Economics and Statistics*, *86*, 25–41.

Binz, M., & Endres, D. (2019). Where do heuristics come from? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41th Annual Meeting of the Cognitive Science Society* (pp. 1402–1408). CogSci 2019: Creativity Cognition Computation.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, *23*(5), 408–422.

Brehmer, B. (1979). Preliminaries to a psychology of inference. *Scandinavian Journal of Psychology*, *20*(4), 193–210.

Brighton, H. (2006). Robust inference with simple cognitive models [Symposium]. *AAAI Spring Symposium: Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems*, 17–22.

Brighton, H., & Gigerenzer, G. (2012). Are rational actor models "rational" outside small worlds? In S. Okasha, & K. Binmore (Eds.), *Evolution and rationality: Decisions, co-operation, and strategic behavior* (pp. 84–109). Cambridge University Press.

Bröder, A. (2000). Assessing the empirical validity of the "take-the-best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1332–1346.

Bröder, A. (2003). Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 611–625.

Bröder, A., & Gaissmaier, W. (2007). Sequential processing of cues in memory-based multiattribute decisions. *Psychonomic Bulletin & Review*, *14*(5), 895–900.

Bröder, A., & Schiffer, S. (2003). Take the best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, *132*(2), 277–293.

Bröder, A., & Schiffer, S. (2006). Stimulus format and working memory in fast and frugal strategy selection. *Journal of Behavioral Decision Making*, *19*(4), 361–380.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *43*(1), 9–64.

Chaitin, G. J. (1969). On the simplicity and speed of programs for computing infinite sets of natural numbers. *Journal of the ACM (JACM)*, *16*(3), 407–422.

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, *90* 1), 63–86.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning phrase representations using RNN encoder-decoder for statistical machine translation*. arXiv preprint arXiv:1406.107

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

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review*, *127*(3), 412–441.

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

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

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

Gauvrit, N., Zenil, H., Delahaye, J.-P., & Soler-Toscano, F. (2014). Algorithmic complexity for short binary strings applied to psychology: A primer. *Behavior Research Methods*, *46*(3), 732–744.

Gauvrit, N., Zenil, H., & Tegnér, J. (2017). The information-theoretic and algorithmic approach to human, animal, and artificial cognition. *Representation and reality in humans, other living organisms and intelligent machines* (pp. 117–139). Springer.

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

Genewein, T., Leibfried, F., Grau-Moya, J., & Braun, D. A. (2015). Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Frontiers in Robotics and AI*, *2*, Article 27.

Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *Cognition*, *204*, Article 104394.

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

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to kahneman and tversky. *Psychological Review*, *103*(3), 592–596.

Gigerenzer, G. (2003). The adaptive toolbox and lifespan development: Common questions? *Understanding human development* (pp. 423–435). Springer.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*(1), 20–29.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143.

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

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

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

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

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

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

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*(3), 227–247.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory*, *9*(6), 408–418.

GPyOpt. (2016). *GPyOpt: A bayesian optimization framework in python*. http://github.com/SheffieldML/GPyOpt

Griffiths, T. L., Daniels, D., Austerweil, J. L., & Tenenbaum, J. B. (2018). Subjective randomness as statistical inference. *Cognitive Psychology*, *103*, 85–109.

Grünwald, P. D., & Grunwald, A. (2007). *The minimum description length principle*. MIT Press.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, *62*(4), 255–262.

Heck, D. W., Hilbig, B. E., & Moshagen, M. (2017). From information processing to decisions: Formalizing and comparing psychologically plausible choice models. *Cognitive Psychology*, *96*, 26–40.

Hilbig, B. E. (2010). Reconsidering "evidence" for fast-and-frugal heuristics. *Psychonomic Bulletin & Review*, *17*(6), 923–930.

Hill, F., Lampinen, A., Schneider, R., Clark, S., Botvinick, M., McClelland, J. L., & Santoro, A. (2020). *Environmental drivers of systematicity and generalization in a situated agent* [Conference session]. International Conference on Learning Representations, Virtual. https://openreview.net/forum?id=SklGryBtwr

Hinton, G. E., & Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In L. Pitt (Eds.), *Proceedings of the sixth annual conference on Computational learning theory* (pp. 5–13). Association for Computing Machinery.

Ho, M. K., Abel, D., Cohen, J. D., Littman, M. L., & Griffiths, T. L. (2020). *The efficiency of human cognition reflects planned information processing* [Conference session]. Proceedings of the 34th AAAI Conference on Artificial Intelligence, New York, United States.

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

Hogarth, R. M., & Karelaia, N. (2006). Take-the-best and other simple strategies: Why and when they work well with binary cues. *Theory and Decision*, *61*(3), 205–249.

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

Honkela, A., & Valpola, H. (2004). Variational learning and bits-back coding: An information-theoretic view to Bayesian learning. *IEEE Transactions on Neural Networks*, *15*(4), 800–810.

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

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*(2), 183–233.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 924–941.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133–156.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

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

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

Katsikopoulos, K. V., Schooler, L. J., & Hertwig, R. (2010). The robust beauty of ordinary information. *Psychological Review*, *117*(4), 1259–1266.

Kingma, D. P., Salimans, T., & Welling, M. (2015). Variational dropout and the local reparameterization trick. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (pp. 2575–2583). Curran Associates, Inc.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, *1*(1), 1–7.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, *37*(6), 1036–1073.

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, *135*(2), 162–183.

Lake, B. M. (2019). Compositional generalization through meta sequence-to-sequence learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (pp. 9788–9798). Curran Associates, Inc.

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

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Lichtenberg, J. M., & Şimşek, Ö. (2017). Simple regression models. *Imperfect Decision Makers: Admitting Real-World Rationality*, *58*, 13–25.

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

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, Article e1.

Lieder, F., Krueger, P. M., & Griffiths, T. (2017). An automatic method for discovering rational heuristics for risky choice. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. CogSci.

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLOS Computational Biology*, *14*(4), Article e1006043.

Luan, S., Schooler, L. J., & Gigerenzer, G. (2014). From perception to preference and on to inference: An approach–avoidance analysis of thresholds. *Psychological Review*, *121*(3), 501–525.

Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing*, *11*(2), 103–121.

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, *118*(3), 393–437.

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

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

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

Mata, R., von Helversen, B., & Rieskamp, J. (2010). Learning to choose: Cognitive aging and strategy selection learning in decision making. *Psychology and Aging*, *25*(2), 299–309.

McAllester, D. (2013). A PAC-bayesian tutorial with a dropout bound. *arXiv preprint arXiv:1307.2118*.

Mikulik, V., Delétang, G., McGrath, T., Genewein, T., Martic, M., Legg, S., & Ortega, P. A. (2020). Meta-trained agents implement bayes-optimal agents. *arXiv preprint arXiv:2010.11223*.

Mockus, J. (1975). On bayesian methods for seeking the extremum. In G. I. Marchuk (Ed.), *Optimization techniques IFIP technical conference* (pp. 400–404). Springer.

Molchanov, D., Ashukha, A., & Vetrov, D. (2017). Variational dropout sparsifies deep neural networks. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (pp. 2498–2507). PMLR.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review*, *14*(3), 505–511.

Newell, B. R., & Lee, M. D. (2011). The right tool for the job? comparing an evidence accumulation and a naive strategy selection model of decision making. *Journal of Behavioral Decision Making*, *24*(5), 456–481.

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best." *Organizational Behavior and Human Decision Processes*, *91*(1), 82–96.

Newell, B. R., Weston, N. J., Tunney, R. J., & Shanks, D. R. (2009). The effectiveness of feedback in multiple-cue probability learning. *Quarterly Journal of Experimental Psychology*, *62*(5), 890–908.

Ortega, P. A., Braun, D. A., Dyer, J., Kim, K.-E., & Tishby, N. (2015). Information-theoretic bounded rationality. *arXiv preprint arXiv:1512.06789*.

Ortega, P. A., Wang, J. X., Rowland, M., Genewein, T., Kurth-Nelson, Z., Pascanu, R., Heess, N., Veness, J., Pritzel, A., Sprechmann, P., Jayakumar, S. M., McGrath, T., Miller, K. J., Azar, M. G., Osband, I., Rabinowitz, N. C., György, A., Chiappa, S., Osindero, S. … Legg, S. (2019). Meta-learning of sequential strategies. *arXiv preprint arXiv:1905.03030*.

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

Parpart, P., Schulz, E., Speekenbrink, M., & Love, B. C. (2017). Active learning reveals underlying decision strategies. *bioRxiv*. https://doi.org/10.1101/239558

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

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

Persson, M., & Rieskamp, J. (2009). Inferences from memory: Strategy- and exemplar-based judgment models compared. *Acta Psychologica*, *130*(1), 25–37.

Rabinowitz, N. C. (2019). Meta-learners' learning dynamics are unlike learners'. *arXiv preprint arXiv:1905.01320*.

Reddi, S. J., Kale, S., & Kumar, S. (2019). On the convergence of Adam and beyond. *arXiv preprint arXiv:1904.09237*.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, & The ABC Research Group. (Eds.), *Simple heuristics that make us smart* (pp. 141–167). Oxford University Press.

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

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies—revisited. *NeuroImage*, *84*, 971–985.

Ritter, S., Barrett, D. G., Santoro, A., & Botvinick, M. M. (2017). Cognitive psychology for deep neural networks: A shape bias case study. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (pp. 2940–2949). PMLR.

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

Samuels, R., Stich, S., & Bishop, M. (2012). Ending the rationality wars. *Collected Papers, Volume 2: Knowledge, Rationality, and Morality, 1978–2010* (Vol. 2, pp. 191–223). Oxford University Press.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167

Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., & Lillicrap, T. (2016). Meta-Learning with memory-augmented neural networks. In M. F. Balcan & K. Q. Weinberger (Eds.), *Proceedings of the 33rd International Conference on Machine Learning* (pp. 1842–1850). PMLR.

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*, *120*(1), 39–64.

Scheibehenne, B., & von Helversen, B. (2009). Useful heuristics. *Making essential choices with scant information* (pp. 195–212). Springer.

Schmidhuber, J., Zhao, J., & Wiering, M. (1996). Simple principles of metalearning. *Technical Report IDSIA*, *69*, 1–23.

Schulz, E., Speekenbrink, M., & Meder, B. (2016). Simple trees in complex forests: Growing take the best by approximate bayesian computation. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society, Recognizing and Representing Events*, CogSci.

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

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

Siegelmann, H. T., & Sontag, E. D. (1992). On the computational power of neural nets. *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 440–449). Association for Computing Machinery.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.

Simon, H. A. (1990a). Bounded rationality. *Utility and probability* (pp. 15–18). Springer.

Simon, H. A. (1990b). Invariants of human behavior. *Annual Review of Psychology*, *41*(1), 1–19.

Sims, C. R. (2018). Efficient coding explains the universal law of generalization in human perception. *Science*, *360*(6389), 652–656.

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

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

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 2951–2959). Curran Associates. https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf

Söllner, A., & Bröder, A. (2016). Toolbox or adjustable spanner? a critical comparison of two metaphors for adaptive decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 215–237.

Solomonoff, R. J. (1964). A formal theory of inductive inference. Part I. *Information and Control*, *7*(1), 1–22.

Szollosi, A., & Newell, B. R. (2020). People as intuitive scientists: Reconsidering statistical explanations of decision making. *Trends in Cognitive Sciences*, *24*(12), 1008–1018. https://doi.org/10.1016/j.tics.2020.09.005

Thrun, S., & Pratt, L. (1998). Learning to learn: Introduction and overview. *Learning to learn* (pp. 3–17). Springer.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, *1*, 211–244.

Todd, P. M., & Dieckmann, A. (2005). Heuristics for ordering cue search in decision making. In L. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (pp. 1393–1400). MIT Press.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*(5), 742–780.

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, *16*(3), 167–171.

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

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

van Ravenzwaaij, D., Moore, C. P., Lee, M. D., & Newell, B. R. (2014). A hierarchical bayesian modeling approach to searching and stopping in multi-attribute judgment. *Cognitive Science*, *38*(7), 1384–1405.

van Rooij, I., Wright, C. D., & Wareham, T. (2012). Intractability and the use of heuristics in psychological explanations. *Synthese*, *187*(2), 471–487.

von Helversen, B., Karlsson, L., Mata, R., & Wilke, A. (2013). Why does cue polarity information provide benefits in inference problems? The role of strategy selection and knowledge of cue importance. *Acta Psychologica*, *144*(1), 73–82.

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.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*(2), 222–236.

Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., Blundell, C., Kumaran, D., & Botvinick, M. (2016). Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*.

Wichary, S., Mata, R., & Rieskamp, J. (2016). Probabilistic inferences under emotional stress: How arousal affects decision processes. *Journal of Behavioral Decision Making*, *29*(5), 525–538.

Yin, M., Tucker, G., Zhou, M., Levine, S., & Finn, C. (2020). Meta-learning without memorization [Conference session]. *International*

Conference on Learning Representations, Virtual. https://openreview.net/forum?id=BklEFpEYwS

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942.

Zenil, H., Marshall, J. A., & Tegnér, J. (2015). Approximations of algorithmic and structural complexity validate cognitive-behavioural experimental results. *arXiv preprint arXiv:1509.06338*.

Zenon, A., Solopchuk, O., & Pezzulo, G. (2019). An information-theoretic perspective on the costs of cognition. *Neuropsychologia*, *123*, 5–18.

Appendix A

Power Analysis

Environments with continuous features can facilitate statistical analysis as fewer trials are needed to observe expected effects. To verify this hypothesis, we conducted a power analysis for an environment with continuous features and one for an environment, where features are dichotomized based on their median. The results presented here are based on an environment with known feature rankings and $T = 10$ decisions per task.

In both settings, we computed how many tasks are on average required to distinguish the single cue heuristic (SC) from the ideal observer model (IO), assuming that decisions are made by the single cue heuristic. In dichotomized environments, ties between features of two options are likely, and hence we modified the single cue heuristic to make decisions based on the first feature that discriminates between both options. We assumed that decisions are made by the single cue heuristic and measured the average support for the single cue heuristic over the ideal observer model on a single task by computing log-Bayes Factors (Kass & Raftery, 1995) between both strategies, see Equations A1 and A2:

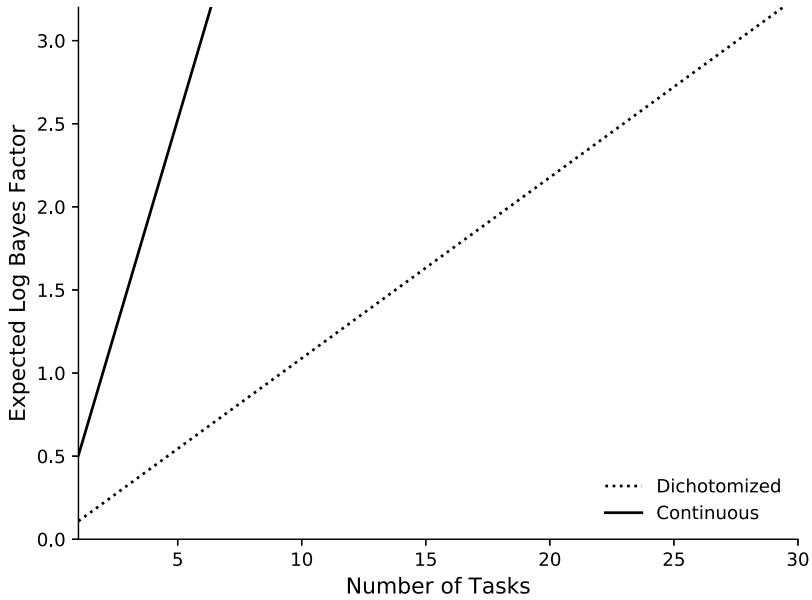
$$\log \mathrm{BF} = \mathbb{E}_{p(\mathbf{x}_{1:T}, y_{1:T})}\left[\sum_{t=1}^{T}\left[\int p(c_t|\mathbf{x}_t, w, m = \mathrm{SC}) \log\left(\frac{p(c_t|\mathbf{x}_t, w, m = \mathrm{SC})}{p(c_t|\mathbf{x}_t, \mathbf{w}, m = \mathrm{IO})}\right) dc_t\right]\right] \tag{A1}$$

$$= \mathbb{E}_{p(\mathbf{x}_{1:T}, y_{1:T})}\left[\sum_{t=1}^{T} \mathrm{KL}[p(c_t|\mathbf{x}_t, w, m = \mathrm{SC})||p(c_t|\mathbf{x}_t, \mathbf{w}, m = \mathrm{IO})]\right]. \tag{A2}$$

The expectation over tasks was approximated using 10^5 samples. Furthermore, we assumed that tasks are sampled independently from each other, meaning that we can multiply log BF by the total number of encountered tasks K to get expected log-Bayes Factors for an experiment with K tasks. Figure A1 shows this analysis for both continuous and dichotomized environments. We observed that it requires roughly four times more tasks to distinguish the single cue heuristic from an ideal observer model in environments with dichotomized features compared to one with continuous features.

Figure A1

Power Analysis for Environments With Known Feature Ranking



Note. The plot illustrates how many tasks are on average required to distinguish the ideal observer model from the single cue heuristic, assuming that decisions are made by the single cue heuristic. We show results for both dichotomized environments (dotted) and environments with continuous features (solid).

(*Appendices continue*)

Appendix B

Variational Inference Details

We update posterior distributions over weights after each observation using variational inference. The true posterior is approximated with a normal distribution $q(\mathbf{w};\lambda_t) = \mathcal{N}(\mathbf{w};\mu_t,\Psi_t)$ and its parameters $\lambda_t = (\mu_t, \mathbf{\Psi}_t)$ are obtained through maximizing the evidence lower bound:

$$\mathcal{L}(\lambda_t) = \mathbb{E}_{\mathbf{w}\sim q(\mathbf{w};\lambda_t)}[\log p(C_t = c_t|\mathbf{x}_t,\mathbf{w})] - \mathrm{KL}[q(\mathbf{w};\lambda_t)||q(\mathbf{w};\lambda_{t-1})]. \tag{B1}$$

The initial prior is set to a standard normal distribution $q(\mathbf{w};\lambda_0) = \mathcal{N}(0,\mathbf{I})$. We furthermore employ a mean-field approximation, in which posterior covariance matrices $\mathbf{\Psi}_t$ are restricted to be diagonal. To ensure positive semidefinite covariance matrices, we parametrize them with logarithms of their standard deviations.

Equation B1 is maximized through gradient-based optimization using AmsGrad (Reddi et al., 2019) with a learning rate of 0.1. Training is stopped once the evidence lower bound function does not increase anymore over 10 steps or after 1,000 total gradient steps.

The Kullback–Leibler divergence can be evaluated in closed-form assuming normal prior and posterior distributions. The expected log-likelihood term is approximated through 100 samples and we employ the reparametrization trick (Kingma & Welling, 2013) to obtain gradients with respect to the variational parameters λ_t.

Appendix C

Meta-Learning Details

Architecture

The architecture of meta-learned inference (MI) and bounded meta-learned inference (BMI) consists of a gated recurrent unit (GRU, Cho et al., 2014) with a hidden size of 128 units, followed by two linear transformations projecting to μ_t and $\log \sigma_t$, respectively. The latter are used to construct diagonal posterior covariance matrices $\mathbf{\Psi}_t$ as in the ideal observer model. The exact forward pass equations are given by Equations C1, C2, C3, C4, C5 and C6:

$$\mathbf{r}_t = \sigma(\mathbf{W}_{ir}[\mathbf{x}_t, c_t] + \mathbf{W}_{hr}\mathbf{h}_{t-1}) \tag{C1}$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_{iz}[\mathbf{x}_t, c_t] + \mathbf{W}_{hz}\mathbf{h}_{t-1}) \tag{C2}$$

$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tanh(\mathbf{W}_{ih}[\mathbf{x}_t, c_t] + \mathbf{W}_{hh}(\mathbf{r}_t \odot \mathbf{h}_{t-1})) \tag{C3}$$

$$\mu_t = \mathbf{W}_{\mu}\mathbf{h}_t \tag{C4}$$

$$\log \sigma_t = \mathbf{W}_{\sigma}\mathbf{h}_t \tag{C5}$$

$$\mathbf{\Psi}_t = \mathrm{diag}(e^{\log \sigma_t}), \tag{C6}$$

where σ denotes the logistic sigmoid function and $\odot$ element-wise multiplication. Together, we denote the set of all model parameters as $\mathbf{\Theta} = \{\mathbf{W}_{ir}, \mathbf{W}_{hr}, \mathbf{W}_{iz}, \mathbf{W}_{hz}, \mathbf{W}_{ih}, \mathbf{W}_{hh}, \mathbf{W}_{\boldsymbol{\mu}}, \mathbf{W}_{\boldsymbol{\sigma}}\}$.

Meta-Learning

MI and BMI are obtained by minimizing Equation 8 with the AmsGrad optimizer (Reddi et al., 2019). During meta-learning, the expectation of the log-likelihood term is approximated through one sample from the encoding distribution $q(\mathbf{\Theta}; \mathbf{\Lambda})$ and we obtain gradients with respect to $\mathbf{\Lambda}$ using the reparametrization trick (Kingma & Welling, 2013). The following pseudocode describes the meta-learning procedure:

(*Appendices continue*)

Algorithm 1: Meta-Learning

while *not converged* **do**

sample a batch of tasks: $\mathbf{x}_{1:T}, c_{1:T} \sim p(\mathbf{x}_{1:T}, c_{1:T})$;

sample model parameters: $\boldsymbol{\Theta} \sim q(\boldsymbol{\Theta}; \boldsymbol{\Lambda})$;

initialize loss: $\mathcal{L}(\boldsymbol{\Lambda}) \leftarrow \beta \mathrm{KL}\left[q(\boldsymbol{\Theta}; \boldsymbol{\Lambda}) || p(\boldsymbol{\Theta})\right]$;

for $t \leftarrow 0$ **to** $T-1$ **do**

compute $\boldsymbol{\lambda}_t = \{\boldsymbol{\mu}_t, \boldsymbol{\Psi}_t\}$ according to Equations C1 to C6;

compute $p(C_{t+1} = 1 | \mathbf{x}_{t+1}, \boldsymbol{\lambda}_t, \boldsymbol{\Theta})$ according to Equation 4;

accumulate loss: $\mathcal{L}(\boldsymbol{\Lambda}) \leftarrow \mathcal{L}(\boldsymbol{\Lambda}) - \log p(C_{t+1} = c_{t+1} | \mathbf{x}_{t+1}, \boldsymbol{\lambda}_t, \boldsymbol{\Theta})$;

end

perform gradient step: $\boldsymbol{\Lambda} \leftarrow \textsc{AmsGrad}(\mathcal{L}(\boldsymbol{\Lambda}), \boldsymbol{\Lambda})$;

end

Learning rates are set to 3×10^{-4} and we train for 10^6 iterations with a batch size of 32; at the end of meta-learning, the loss function has converged. Each model is initialized from a pretrained version without resource limitations and we increase β linearly over the first half of the training to the desired value.

Evaluation

During evaluation the expectation of the log-likelihood term is approximated through $K = 100$ samples from the encoding distribution and we perform no further updates of meta-parameters:

Algorithm 2: Evaluation

Input: a particular task $\mathbf{x}_{1:T}, c_{1:T}$

sample model parameters: $\boldsymbol{\Theta}_k \sim q(\boldsymbol{\Theta}; \boldsymbol{\Lambda})$;

for $t \leftarrow 0$ **to** $T-1$ **do**

for $k \leftarrow 1$ **to** K **do**

compute $\boldsymbol{\lambda}_{t,k} = \{\boldsymbol{\mu}_{t,k}, \boldsymbol{\Psi}_{t,k}\}$ according to Equations C1 to C6;

compute $p(C_{t+1} = 1 | \mathbf{x}_{t+1}, \boldsymbol{\lambda}_{t,k}, \boldsymbol{\Theta}_k)$ according to Equation 4;

end

compute predictive posterior distribution: $\frac{1}{K} \sum_{k=1}^{K} p(C_{t+1} = 1 | \mathbf{x}_{t+1}, \boldsymbol{\lambda}_{t,k}, \boldsymbol{\Theta}_k)$;

end

(*Appendices continue*)

Prior

The prior over meta-parameters corresponds to a variational dropout prior (Kingma et al., 2015). In variational dropout, model parameters are corrupted by multiplicative normally distributed noise, see Equations C7, C8 and C9:

$$\boldsymbol{\Theta}_i = \mu_i \cdot \xi_i \tag{C7}$$

$$\xi_i \sim \mathcal{N}(\xi_i|1, \alpha_i) \tag{C8}$$

$$\Rightarrow q(\boldsymbol{\Theta}; \boldsymbol{\Lambda}) = \prod_i \mathcal{N}(\boldsymbol{\Theta}_i|\mu_i, \alpha_i\mu_i^2). \tag{C9}$$

Instead of parametrizing the encoding distribution by $\boldsymbol{\Lambda} = \{\boldsymbol{\mu}_i, \boldsymbol{\alpha}_i\}$, Molchanov et al. (2017) suggested the following reparametrization (Equations C10 and C11) to reduce the variance of stochastic gradients:

$$\sigma_i^2 = \alpha_i \cdot \mu_i^2 \tag{C10}$$

$$\Rightarrow q(\boldsymbol{\Theta}; \boldsymbol{\Lambda}) = \prod_i \mathcal{N}(\boldsymbol{\Theta}_i|\mu_i, \sigma_i^2), \tag{C11}$$

which is used together with an improper log-scale uniform prior over model parameters as described in Equation C12:

$$p(|\boldsymbol{\Theta}_i|) \propto \frac{1}{|\boldsymbol{\Theta}_i|}. \tag{C12}$$

There is no analytical expression for the KL term (ref. Equation 8) under this prior and encoding distribution, however it can be approximated numerically. Molchanov et al. (2017) suggested the following approximation:

$$\mathrm{KL}[q(\boldsymbol{\Theta}_i|\boldsymbol{\Lambda}_i)||p(\boldsymbol{\Theta}_i)] \approx -k_1\sigma(k_2 + k_3 \log \alpha_i) + 0.5 \log(1 + \alpha_i^{-1}) - \text{const.} \tag{C13}$$

$$k_1 = 0.63576, k_2 = 1.87320, k_3 = 1.48695. \tag{C14}$$

Meta-learning models presented in this article use the parametrization from Equation C11 and approximate the KL term through Equations C13 and C14.

Appendix D

Bayesian Model Comparison

We relied on Bayesian model comparisons (Bishop, 2006) to test which hypothesis accounted best for human choices. For the most part, we performed separate comparisons for each participant in order to detect potential individual differences.

Let $\mathbf{X}^{(i)} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{KT}\}$ denote the set of all observed features and $\hat{\mathbf{c}}^{(i)} = \{\hat{c}_1, \ldots, \hat{c}_{KT}\}$ the set of corresponding decisions from a single participant i, and let $\mathbf{X}$ and $\hat{\mathbf{c}}$ denote the joint data for all participants. K corresponds to the total number of tasks and T to the number of trials per task. Note that we use $\hat{c}$ to refer to decisions made by participants and c to refer to ground truth labels. We can then compute the probability that a participant used strategy m through Bayes' rule as described in Equation D1:

$$p(m|\hat{\mathbf{c}}^{(i)}, \mathbf{X}^{(i)}) = \frac{p(\hat{\mathbf{c}}^{(i)}|\mathbf{X}^{(i)}, m)p(m)}{p(\hat{\mathbf{c}}^{(i)}|\mathbf{X}^{(i)})}. \tag{D1}$$

We assumed a uniform prior over hypothesis in all of our analyses. For models that include fitted parameters, we approximated the model evidence using the Bayesian information criterion (BIC, Equation D2, Schwarz, 1978):

$$\log p(\hat{\mathbf{c}}^{(i)}|\mathbf{X}^{(i)}, m) \approx -\frac{1}{2}|\theta| \log \mathrm{KT} + \log p(\hat{\mathbf{c}}^{(i)}|\mathbf{X}^{(i)}, \theta^*, m) + \text{const.} \tag{D2}$$

where $|\theta|$ denotes the number of parameters and θ^* a maximum likelihood estimate. For all models except BMI, the maximum likelihood estimate was obtained using Bayesian optimization (GPyOpt, 2016; Mockus, 1975; Snoek et al., 2012). For BMI, we instead adopted a simple grid-search procedure. Fitted parameters and their search domains were:

Model	Parameter	Domain
Ideal observer	σ	[0.01, 10]
Equal weighting	σ	[0.01, 10]
Single cue	σ	[0.01, 10]
Strategy selection	σ	[0.01, 10]
Feedforward network	σ	[0.01, 10]
	α	[0, 0.1]
Bounded meta-learned inference (BMI)	β	{0, 0.0003, 0.001, 0.003, 0.01, 0.03, 0.1}

(*Appendices continue*)

Appendix E

Strategy Selection Model

Our strategy selection model is based on the idea of Bayesian model selection (Bishop, 2006). In time-Step $t + 1$, the agent selects the model m with the highest posterior probability given the past data $p(m|x_{1:t}, c_{1:t})$ from a set of candidate models $\mathcal{M}$. In our model simulations, we defined the set of candidate models as $\mathcal{M} = \{\text{IO}, \text{SC}, \text{EW}\}$ and assume a uniform prior over models. The computation of the posterior distribution over models can be expressed iteratively:

$$m_{t+1} = \arg\max_{m \in \mathcal{M}} [\log p(m|\mathbf{x}_{1:t}, c_{1:t})] \tag{E1}$$

$$= \arg\max_{m \in \mathcal{M}} [\log p(c_{1:t}|\mathbf{x}_{1:t}, m) + \log p(m)] \tag{E2}$$

$$= \arg\max_{m \in \mathcal{M}} [\log p(c_{1:t}|\mathbf{x}_{1:t}, m)] \tag{E3}$$

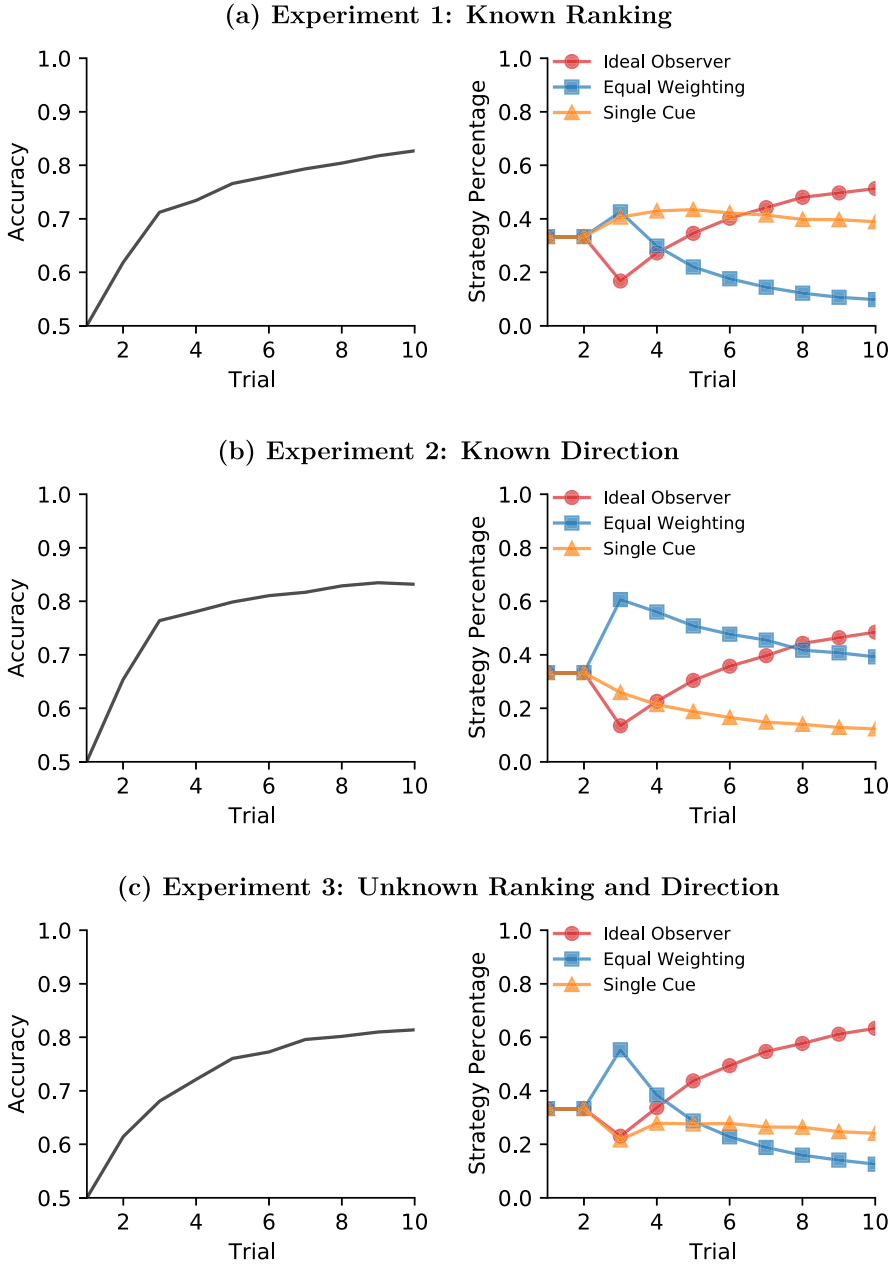
$$= \arg\max_{m \in \mathcal{M}} [\log p(c_t|\mathbf{x}_t, \mathbf{x}_{1:t-1}, c_{1:t-1}, m) + \log p(c_{1:t-1}|\mathbf{x}_{1:t-1}, m)], \tag{E4}$$

Equations E1, E2, E3 and E4 reveal that this strategy selection model amounts to selecting the model with the highest accumulated log evidence over all previous time-steps. The strategy selection model combines advantages of the ideal observer model with those of heuristics: If additional information is provided heuristics may outperform the ideal observer early on and hence they will be initially preferred. However, after a while, the ideal observer model surpasses both heuristics in terms of performance and hence it will be preferred during the later stages of a task. Figure E1 shows the average performance of the strategy selection model and its probability for the selection of each strategy plotted over the number of trials.

(*Appendices continue*)

Figure E1

Percentage of Correct Decisions (Left) and Probability of Selecting Each Strategy (Right) in the Strategy Selection Model Plotted Over Number of Trials



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

(*Appendices continue*)

Appendix F

Feedforward Network

Our feedforward neural network models use the same architecture as MI and BMI, but without recurrent connections and the previous target as additional input. Parameter updating is performed through gradient descent on the negative log-likelihoods of targets. Figure F1 shows the average performance of the feedforward neural network with different learning rates together with the Gini coefficients of its inferred weight vectors. In our model comparisons, we treated the learning rate α as a free parameter that is fitted to the empirical data. The exact forward pass equations are given by Equations F1, F2, F3, F4, F5 and F6:

$$\mathbf{r}_t = \sigma(\mathbf{W}_{ir}\mathbf{x}_t) \tag{F1}$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_{iz}\mathbf{x}_t) \tag{F2}$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \tanh(\mathbf{W}_{ih}\mathbf{x}_t + \mathbf{r}_t) \tag{F3}$$

$$\mu_t = \mathbf{W}_{\mu}\mathbf{h}_t \tag{F4}$$

$$\log \sigma_t = \mathbf{W}_{\sigma}\mathbf{h}_t \tag{F5}$$

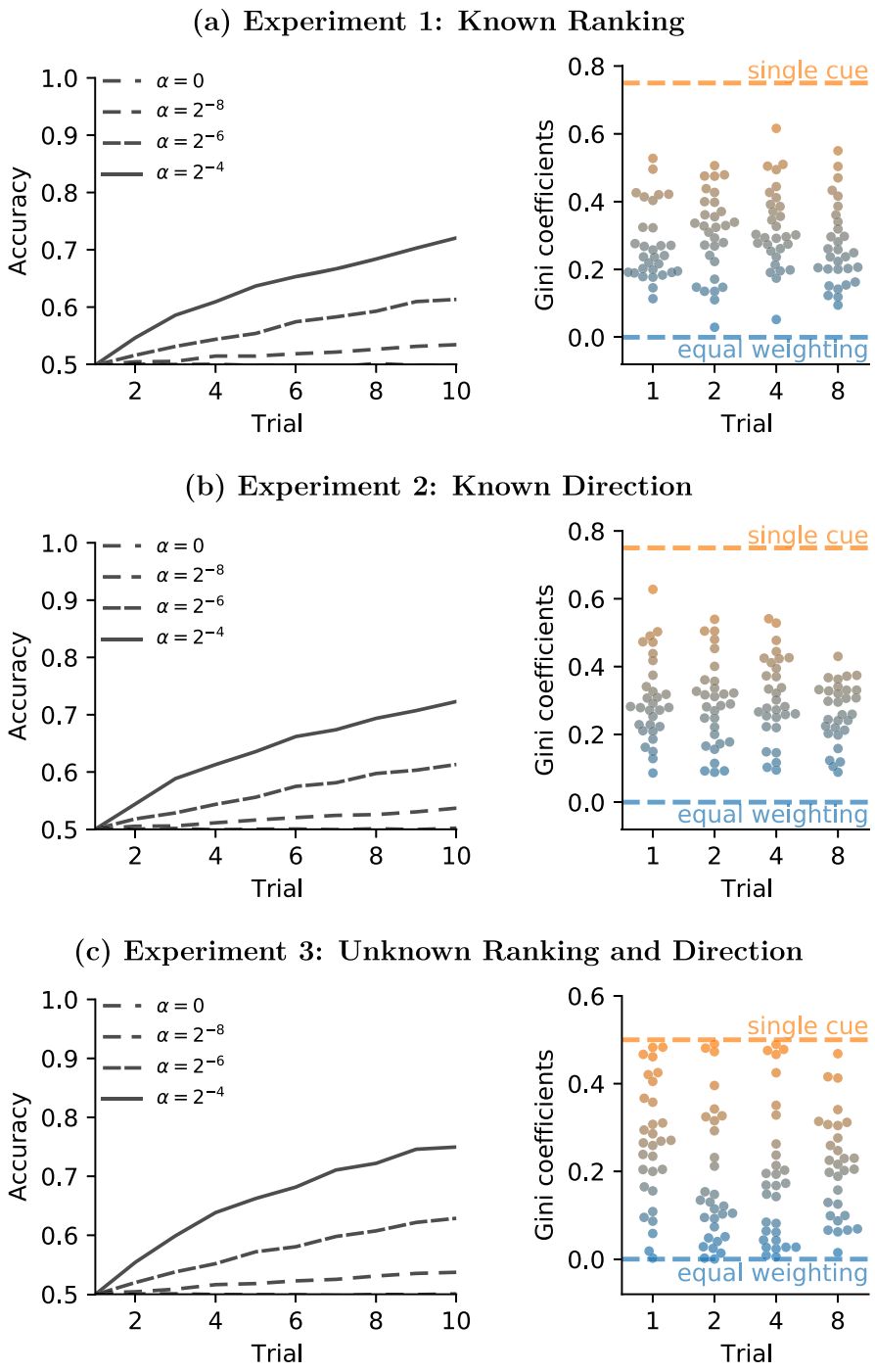
$$\mathbf{\Psi}_t = \mathrm{diag}(e^{\log \sigma_t}), \tag{F6}$$

where σ denotes the logistic sigmoid function and $\odot$ element-wise multiplication.

(*Appendices continue*)

Figure F1

Percentage of Correct Decisions (Left) and Gini Coefficients (Right) for the Feedforward Neural Network Plotted Over Number of Trials



Note. Gini coefficients are shown for an example model with learning rate of 2^{-4} but are similar for other learning rates. See the online article for the color version of this figure.

(*Appendices continue*)

Appendix G

Experiment 3b: Unknown Ranking and Direction, Four Features

Here, we briefly summarize the results of our original study with four features and no information about ranking and direction. We have excluded this study from the main text because the performance of participants did not allow us to draw decisive conclusions about their use of strategies. The design was identical to Experiment 3, except that participants observed four features per alien.

Participants

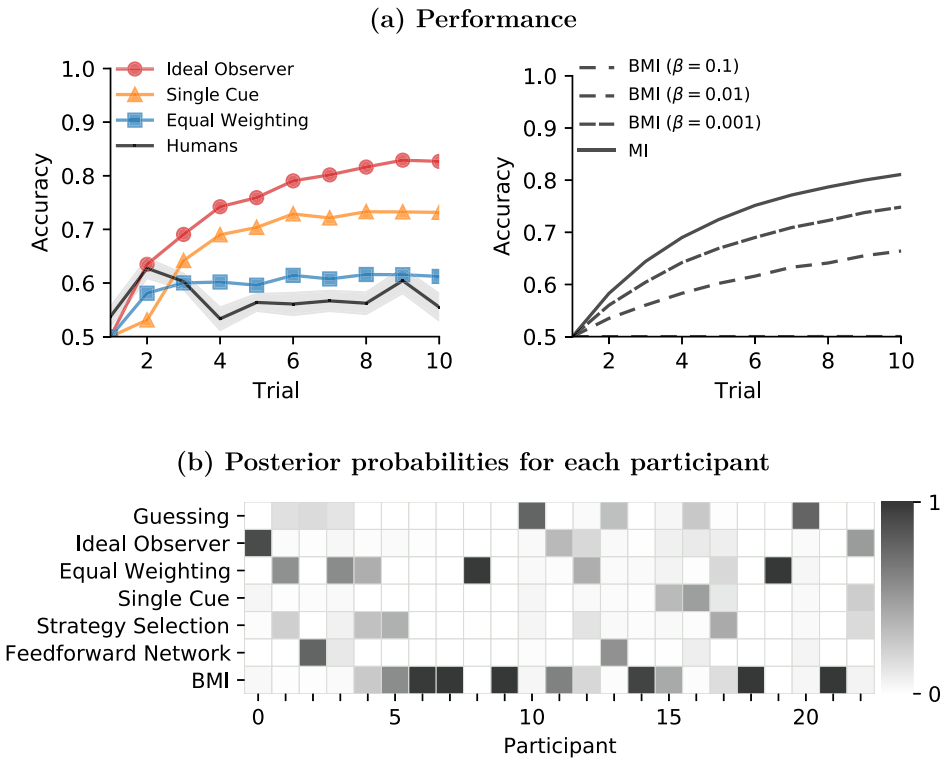
Participants were students from the University of Marburg, taking part in the study for course credits. Besides course credits, they got a chance to win a €10 voucher if they made more than 60% correct decisions. The experiment was approved by the local ethics board (AZ 2020-32k). In total, we collected data from 23 participants (16 female, average age: 23.09 ± 4.38). The median time to complete the experiment was 36.09 min.

Performance

Participants found this version much harder and performed substantially worse. Without the additional information from the first two conditions, their cognitive resource limitations became a dominating factor. The average performance dropped to 57.14% ± 4.38%, ref. Figure G1(a). While some participants performed well, a substantial amount was at or close to chance level. We used an exact binomial test with a base probability of $p = .5$ to assess whether or not individual participants chose the better option more frequently than chance. In this study, only 14 out of 23 participants performed better chance. We then repeated the mixed-effects logistic regression analysis described in the main text to investigate participants' learning over trials and tasks. The results of this model showed no significant fixed effect of either trial number ($\beta = -0.01$, $z = -0.27$, $p = .79$) or task number ($\beta = -0.01$, $z = -0.29$, $p = .77$) onto choosing the better option.

Figure G1

Study Results With Unknown Ranking and Direction, Four Features



Note. BMI = bounded meta-learned inference. (a) Percentage of correct decisions in the unrestricted condition plotted over the number of trials within a task. For human performance shaded contours represent the standard error of the mean. The left panel shows the ideal observer model and both heuristics, while the right shows BMI for different values of β. For BMI, lower β-values correspond to a less restricted model. (b) Posterior distributions for each participant over different strategies in the unrestricted condition. See the online article for the color version of this figure.

(*Appendices continue*)

Model Comparison

Posterior probabilities obtained from a Bayesian model comparison in Figure G1(b) indicated a trend toward strategies that combine information from multiple features. Nine out of 23 participants were best described by BMI; in four of those we found decisive evidence $(p(m = \text{BMI}|\hat{\mathbf{c}}^{(i)}, \mathbf{X}^{(i)}) = 0.99)$. Amongst the participants not best described by BMI, six were best described by the equal weighting heuristic, two by guessing, two by the ideal observer model, two by the feedforward network, one by the single cue heuristic, and one by the strategy selection model. The protected exceedance probability (PXP) provided moderate support for the hypothesis that BMI was the most frequent explanation for participants in our population (PXP = 0.57). The obtained evidence was however not decisive.

Received January 12, 2021
Revision received August 16, 2021
Accepted August 23, 2021 ■