Ecological Rationality

Fast and frugal heuristics can not only perform as well as more complex algorithms, they can also perform better. Even if humans and other animals had the computational resources to use such complex algorithms, following them might be inefficient and result in poor performance. The surprisingly high performance of heuristics results from their ecological rationality. A heuristic is ecologically rational when it performs well by exploiting the statistical structure of the environment. We study ecological rationality from several perspectives, three of which we will focus on in this section. First, we use computer simulation to test fast and frugal heuristics in different environments. Second, we derive analytic results which specify the conditions under which a simple heuristic can match or outperform more complex strategies. Finally, we explore whether people use heuristics, and how they adapt to different environments by selecting different heuristics from the adaptive toolbox.

Simple Heuristics for Inductive Inference

Humans and other animals need to make inductive inferences. This is the task of making an informed guess about the future using observations of the past. An influential trend in contemporary psychology is to examine human performance from the perspective of rational principles of induction, such as Bayesian statistics. This poses a puzzle. Although humans appear to be extremely effective at making inductive inferences, and sometimes act as if they followed rational principles, decades of research into machine learning and pattern recognition have shown that classically rational induction becomes intractable, from a mechanistic perspective, for anything but trivial problems (Brighton & Todd, 2006). How can humans perform so effectively despite constraints on speed, computational resources, and the availability of information?

We examine how simple cognitive mechanisms adapted to natural contexts can shed light on this puzzle. Previously, we have shown that the simple heuristic Take The Best often makes more accurate predictions than sophisticated linear regression models (Czerlinski et al., 1999). Many found these results surprising, but others proposed putting our models to a far tougher test. Addressing this challenge, we compared Take The Best with several resource intensive and sophisticated nonlinear models. These models are commonplace in research into artificial intelligence as well as cognitive modeling. Figure 2 outlines three classic approaches to nonlinear processing: connectionist models, exemplar models, and decision tree models.

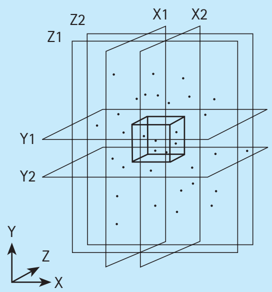
How well does Take The Best compare to the more resource intensive connectionist, exemplar, and rule-based methods? In particular, by performing *less* processing, will performance suffer? For a number of environmental settings, we performed a model comparison to assess the predictive ability of Take The Best and several nonlinear models. In a learning phase, each model is first presented with partial experience of the environment. We then examine how well each model predicts future events in the same environment. Examples of environments include the population of German cities where the task is to infer which of two novel cities has a greater population, or the salaries of professors, where the task is to infer which of two professors earns more. Brighton (2006) demonstrated that Take The Best, even though it performs less processing, can outperform five well-known connectionist, exemplar, and rule-based models of inductive inference.

For three environments, the impressive performance of Take The Best is shown in Figure 3, which plots the predictive accuracy of Take The Best and three competing models as a function of degree of exposure to the environment. Take The Best outperforms the other models by a significant margin. These results strengthen our previous findings which suggested that inductive inference does not demand intensive and domain-general processing methods in order to achieve high performance. Simple heuristics

Key Reference

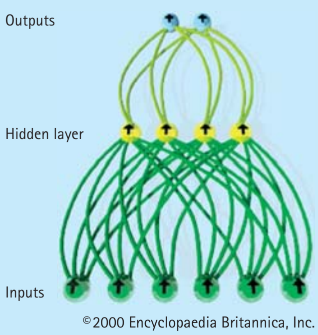
Brighton, H. J. (2006). Robust inference with simple cognitive models. In C. Lebiere & B. Wray (Eds.) *Between a rock and a hard place: Cognitive science principles meet AI-hard problems. Papers from the AAAI Spring Symposium* (AAAI Tech. Rep. No. SS–06–03, pp. 17–22). Menlo Park, CA: AAAI Press.

(A) Exemplar Models



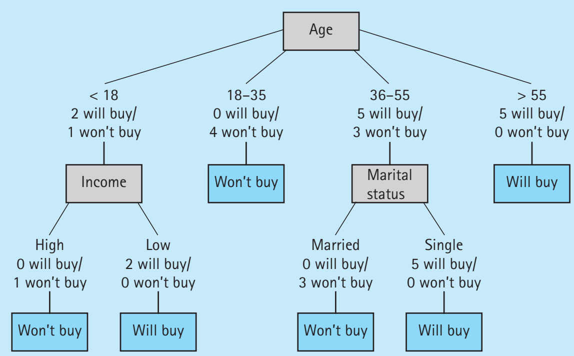
Exemplar models use the metaphor of the memory system, and simply store past observations. These observations are represented as points in a multidimensional feature space. When a novel problem is encountered, the most relevant past observations are retrieved and used to make a prediction.

(B) Neural Network Models



Neural network models use the metaphor of the brain to inform the processing problems by implementing networks of artificial neurons. During learning, the strength of the connections between neurons is adjusted to fit past experience. Neural activations resulting from novel problems are then used to derive a prediction.

(C) Decision Tree Models



Decision tree models construct hierarchies of if-then rules to describe and categorize past experiences. These rules define the patterns that are used to make inferences about novel problems.

Figure 2. Learning algorithms simulate the task of making inductive inferences. They process sequences of observations in an attempt to uncover predictive patterns. Once found, these patterns are then used to predict future events. We evaluated the simple heuristic Take The Best by examining how well it predicts future events in comparison to a number of resource-intensive processing models drawn from three well-known paradigms: (A) exemplar/nearest neighbor models, which store past experiences verbatim, and then retrieve relevant ones them when a confronted with a new problem; (B) feed-forward neural network models, which encode past experiences by learning activation strengths between simulated neurons that map inputs to outputs; (C) decision tree induction models which construct hierarchies of IF-THEN rules to categorize past observations.

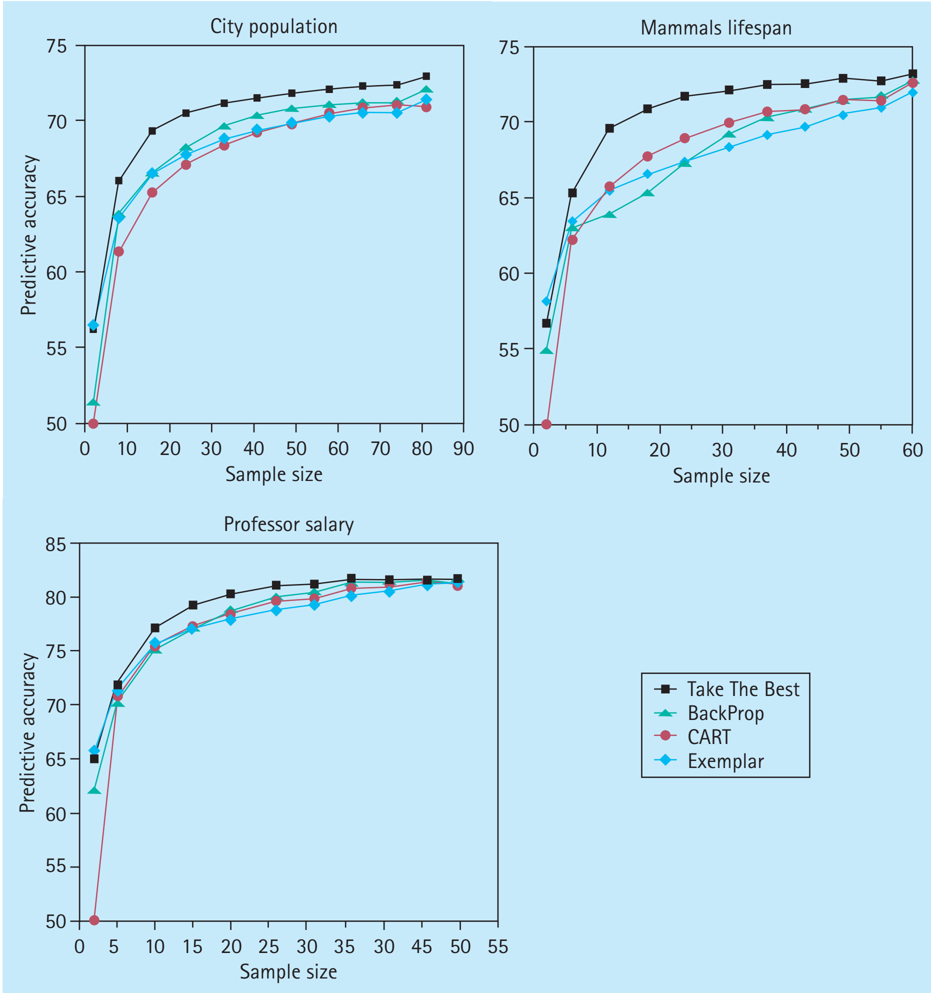


Figure 3. These plots show how the predictive accuracy of the simple heuristic Take The Best compares to three standard machine learning algorithms, in three environments. The rival algorithms include a decision tree induction model (CART), an exemplar model (Exemplar), and a neural network model (BackProp). As a result of performing less processing, Take The Best can often significantly outperform these standard models.

selected from the mind's adaptive toolbox, which fit the problem at hand, do just as well or better. This is a clear and striking example of how simple processes, by being ecologically rational, can lead to adaptive behavior.

When Do Simple Heuristics Perform As Well As a Naive Bayesian?

Simple heuristics can perform as well or better than more resource-intensive models when observations are limited, but can simple heuristics also compete with these models when the decision maker has had complete exposure to the environment? A decision maker who ranks two objects on a criterion (e.g., price) by using several probabilistic cues is *naive* if he or she assumes that cues are independent when the value of the criterion is known. This independence assumption has been made by many authors when examining Bayesian methods for inductive inference (Domingos & Pazzani, 1997; Krauss & Martignon, 2003). Katsikopoulos and Martignon (2006) have studied under which conditions simple heuristics perform as well as a Naive Bayesian model, which assumes independent cues. In particular, they focus on *Take The Best,* which searches through cues in order of their validity, and makes a decision on the basis of a single cue. They also examined *Tally,* which sums up the positive evidence for each object and selects the object with the largest sum. Katsikopoulos and Martignon (2006) showed that Tally is as accurate as Naive Bayes when all cues have the same

Key Reference

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

Key Reference

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

validity. Furthermore, they show that Take The Best performs as well as Naive Bayes when the odds ratio of the cues validities follow a specific structure. The odds ratio is the validity of a cue divided by 1 minus the validity. If this ratio for each cue is larger than the product of the odds ratio for all other cues with lower validities, the *odds condition,* then Take The Best performs as well as Naive Bayes.

These results extend our previous analytic work, studying under which conditions Take The Best is more or equally accurate as Tally, and the converse (Martignon & Hoffrage, 1999, 2002). Among 20 natural environments tested, Katsikopoulos and Martignon (2006) found that the odds condition held in three of them, and speculated that the human mind could be wired to detect the presence of this condition, which, in turn, might trigger the use of Take The Best. In previous work, Krauss and Martignon (2003) found that the proportion of participants that used Take The Best jumped from 20% to 75% in these cases. This work provides a valuable insight into what features of the environment might trigger the selection of one strategy over another.

Selecting Strategies From the Adaptive Toolbox

The adaptive toolbox is the repertoire of specialized cognitive mechanisms, which includes fast and frugal heuristics. One open question is how heuristics are selected from the adaptive toolbox. The idea that people are equipped with a repertoire of cognitive strategies can be found in many areas of psychology, and this makes the strategy selection problem a pressing issue for many researchers. The traditional view follows a cost-benefit approach, where individuals are viewed as trading off the cost of a strategy against its benefits when performing strategy selection. The cost of a strategy is related to the cognitive effort required to execute it, and the benefits are related to its accuracy. According to this view, people *anticipate* both the benefits and costs associated with different strategies, and choose the most cost effective strategy for the problem at hand. This approach has been criticized as not being sufficiently well-specified, making it necessary to advance the theoretical approach by providing a computational model that describes the strategy selection process more precisely.

Rieskamp and Otto (2006) have argued that people do not apply a metastrategy which trades off costs and benefits. If they did, how would they select the metastrategy? To avoid the problem of infinite regress, Rieskamp and Otto (2006) suggested that people select strategies through learning. They aimed to shed light on three fundamental questions: First, do people select different strategies in different environments? Second, do people learn to select the strategy of the adaptive toolbox that performs best in a particular environment? Finally, how can the learning

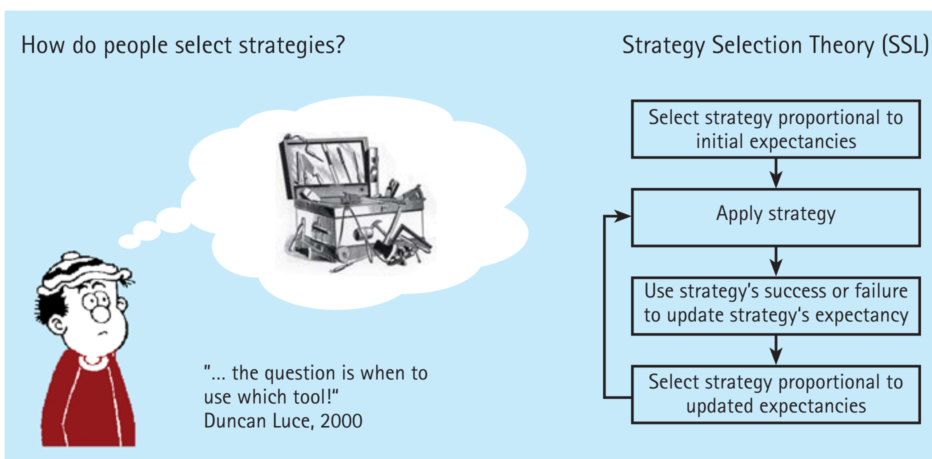


Figure 4. Rieskamp and Otto (2006) consider the problem of how the cognitive system selects between the heuristics in the adaptive toolbox. They propose that these heuristics are undergoing reinforcement. Reinforcement occurs as a result of feedback reflecting how well each heuristic performs in the given context.

process underlying strategy selection best be described? Rieskamp and Otto (2006) focused on the problem of choosing which of two alternatives, described by several probabilistic cues, has a higher criterion value. For instance, which of two companies described by several cues is most creditworthy? When making such an inference, a number of inference strategies could be applied, for example, Take The Best, Tally, or a Weighted Additive strategy. Recent studies have shown that noncompensatory strategies like Take The Best perform well at predicting inferences when participants are under time pressure, when there are high costs in information acquisition, or when information about cues has to be retrieved from memory. In contrast, compensatory strategies like the Weighted Additive strategy were more suitable for making inferences in situations of low time pressure, low information acquisition costs, or when the information was provided simultaneously via a computer screen.

Rieskamp and Otto (2006) proposed a new *Strategy Selection Learning* (SSL) theory, which views strategy selection as a process of reinforcement learning (Figure 4). According to SSL, people use unobservable cognitive strategies, rather than stimulus response associations, and these cognitive strategies are reinforced though feedback. The "expectancies" of the strategies are updated through reinforcement learning, which relies on feedback on past performance. SSL predicts that the strategy that performs best will be selected, given sufficient learning opportunity. Rieskamp and Otto (2006) tested SSL against four alternative models, including three general learning models and one exemplar model. They performed a number of experimental studies in which participants made repeated inferences in different environments, and were given feedback on their performance. In one study, participants had to select which of two unnamed companies was more creditworthy. This experiment considered two environment conditions. In the first condition, featuring a noncompensatory environment, Take The Best yielded the highest performance. In the second condition, featuring a compensatory environment, a Weighted Additive strategy achieved the highest per-

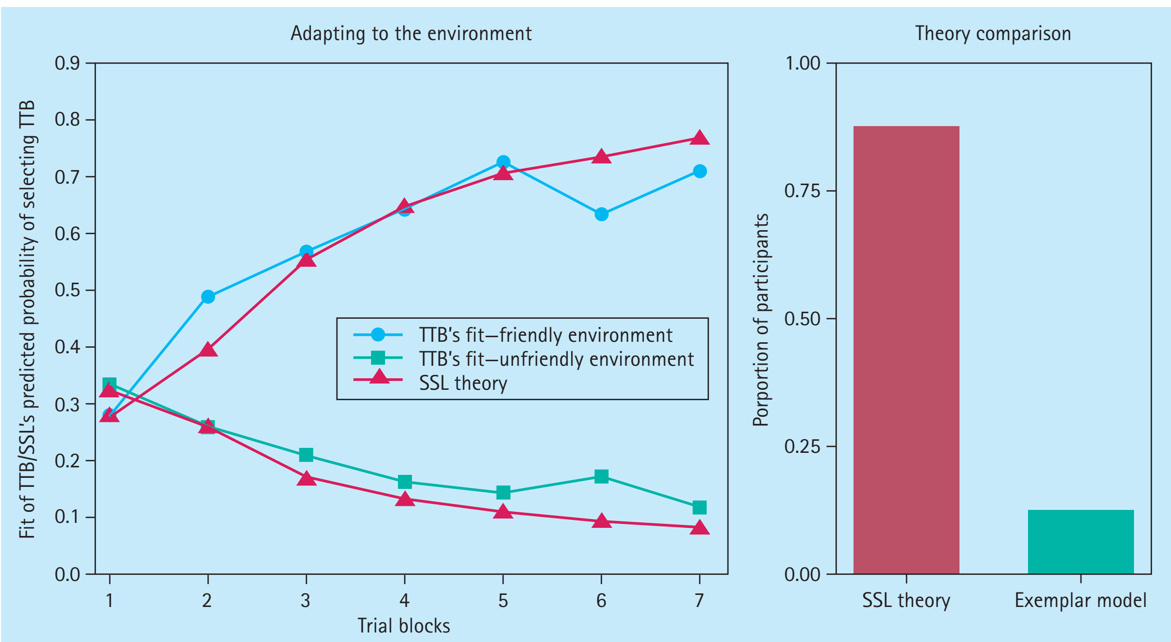


Figure 5. The proportion of choices predicted by the simple heuristic Take The Best increases over time in the Take The Best-friendly environment (left). This adaptation process is predicted by the SSL Theory. The SSL Theory also predicts the behavior of the majority of participants better than a competing exemplar model (right).

Key References

Rieskamp, J. (2006a) Perspectives of probabilistic inference: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1355–1370.

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

formance. This study demonstrated that when people repeatedly make probabilistic inferences, their performance improves. Figure 5 shows that the participants initially had a preference to integrate the provided information, which is not surprising, given that all information was provided and information search was not required. This initial preference changed very quickly depending on the structure of the environment. When the environment structure favored the integration of the available information, people selected the Weighted Additive strategy more frequently. However, when the structure of the environment favored the selection of Take The Best, people selected Take The Best more frequently.

At the end of the experiment, the strategy that performed best in the environment was also the strategy that predicted participants' inferences most accurately. This learning process was also accurately described by the SSL theory (see Figure 5). Interestingly, the more complex learning models and the exemplar model failed to describe the learning process more accurately. This finding was consistent across several inference situations, including a situation involving costly information search, and a situation where a choice between three rather than two alternatives was required. Furthermore, Rieskamp (2006a) compared the SSL theory with a well-established connectionist model, and found further evidence for the adaptive selection of strategies from an adaptive toolbox.

Do Animals Use Simple Heuristics?

For biologists, the concept of ecological rationality is uncontroversial. The assumption that organisms are often well-adapted to their environment is ubiquitous in biology. Consequently, biologists see no reason why this assumption should not be extended to the study of cognition. Furthermore, there is a long tradition in biology of explaining adaptive behavior using *rules of thumb,* which are similar in spirit to the simple heuristics studied by ABC. Hutchinson and Gigerenzer (2005b) explored the similarities and differences between ABC's work and that being carried out in behavioral ecology. The hope was that both disciplines could learn from each other.

One advantage that biologists have over psychologists is that animals can be studied in the natural environments to which they have adapted, although in both disciplines it is often maladaptive behavior in atypical environments that spurs research. Interestingly, the range of animal behavior described using rules of thumb is far broader than the range of human behavior explored by ABC. The rules of thumb used by animals have been discovered largely through careful empirical observation and experiments, and there are real prospects for understanding the neurological and physiological processes underlying them. Mirroring ABC's approach, there are also simulation studies in biology that assess the performance of rules of thumb.

Can biologists learn from ABC's work? The idea that simple rules of thumb can outperform more complex processes has not been thoroughly explored, or widely appreciated, in biology. From ABC's perspective, a rule of thumb is likely to exist as an adaptive response to the ecological structure of the task environment, rather than existing merely as a result of a simple brain. Other aspects lacking in biology are (a) a theoretical perspective capable of explaining the enormous diversity of methods of cue integration biologists have documented and (b) the development of precise models specifying how multiple cues are processed. Although ABC also has much progress to make in explaining why the statistical structure of environments favors particular methods of cue integration, biologists could benefit from our attempts to model and understand this aspect of cognition. Ultimately, the conceptual vocabulary needed to address this fundamental problem is likely to be the same for both disciplines.