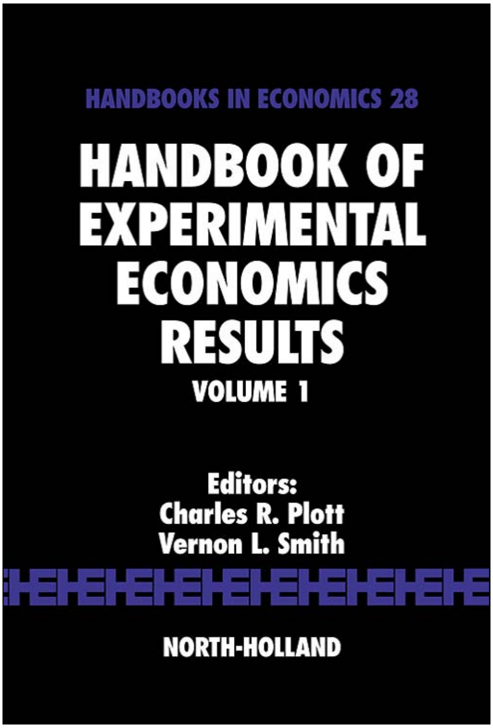

Gerd Gigerenzer, Laura Martignon, Ulrich Hoffrage, Jörg Rieskamp, Jean Czerlinski and Daniel G. Goldstein, One-Reason Decision Making. In: C.R. Plott and V.L. Smitt, editors, Handbook of Experimental Economics Results, Volume 1.
Amsterdam: North-Holland, 2008, p. 1004.
ISBN: 978-0-444-82642-8

Chapter 108

ONE-REASON DECISION MAKING

GERD GIGERENZER, LAURA MARTIGNON, ULRICH HOFFRAGE, JÖRG RIESKAMP and JEAN CZERLINSKI

Max Planck Institute for Human Development, Berlin, Germany

DANIEL G. GOLDSTEIN

Columbia University, New York, USA

"One-reason decision making" is a label for a class of fast and frugal heuristics that base decisions on only one reason. These heuristics do not attempt to optimally fit parameters to a given environment; rather, they have simple structural features and "bet" that the environment will fit them. By not attempting to optimize, these heuristics can save time and computations, and demand only little knowledge concerning a situation. Models of one-reason decision making have been designed for various tasks, including choice, numerical estimation, and classification (Gigerenzer, Todd, and the ABC Research Group, 1999). In this chapter, we focus on two of these heuristics, "Take The Best" and Minimalist, and compare their performance with that of standard statistical strategies that weigh and combine many reasons, such as multiple regression. Contrary to common intuition, more reasons are not always better.

1. "Take The Best" and Minimalist

We deal with two-alternative prediction tasks, such as which of two American cities will have the higher homelessness rate, or which of two stocks will yield a higher return. In general terms, the task is to predict which object, a or b, has the higher value on a criterion. There is a set of N objects and a set of M cues. In the case of binary cues, cue values "1" and "0" indicate higher and lower criterion values, respectively. Take The Best can be characterized by the following building blocks (see also Gigerenzer and Goldstein, 1996):

(0) Recognition heuristic: If only one object is recognized, and recognition is positively correlated with the criterion, predict that this object has the higher value on the criterion. If neither is recognized, then guess. If both are recognized, go on to Step 1.

(1) Search rule: Choose the cue with the highest validity and look up the cue values of the two objects.

(2) Stopping rule: If one object has a cue value of one ("1") and the other does not (i.e., "0" or unknown), then stop search and go on to Step 3. Otherwise exclude this cue and go back to Step 1. If no cues are left, guess.

Handbook of Experimental Economics Results, Volume 1

DOI: 10.1016/S1574-0722(07)00108-4

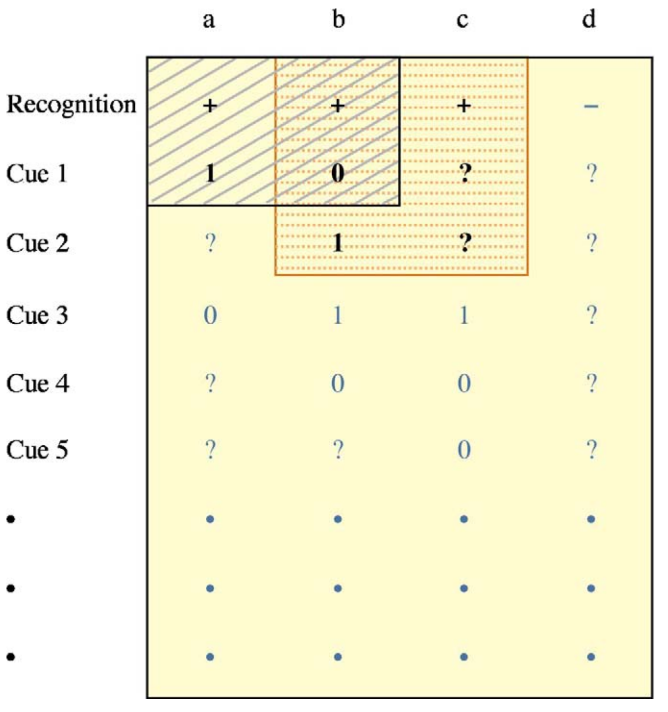

Figure 1. Illustration of one-reason decision making by the Take The Best heuristic. Objects a, b, and c are recognized (+), d is not (−), Cue values are binary, "1" and "0" indicate higher and lower criterion values, respectively. Missing knowledge, that is, unknown cue values are denoted by a question mark. For instance, to infer whether $a > b$, Take The Best looks up only the values in the gray-striped space. To infer whether $b > c$, search is bounded to the dotted space. In each case, the decision is based on only one cue; the cue values of less important cues are not even looked up.

(3) Decision rule: Predict that the object with the cue value of one ("1") has the higher value on the criterion.

The recognition heuristic (Step 0) only plays a role in situations of partial ignorance: when some of the N objects are unknown, such as when one recognizes only a subset of brand names (Goldstein and Gigerenzer, 2002). The validity v_i of a cue i (Step 1) is defined as

$$v_i = \frac{R_i}{R_i + W_i},$$

where R_i is the number of right (correct) inferences, and W_i is the number of wrong (incorrect) inferences based on cue i alone. $R_i + W_i$ equals the number of cases where one object has the value "1" and the other does not.

Figure 1 illustrates the logic of Take The Best. Search for information is stopped when the first cue is found on which the two alternatives differ. This stopping rule does

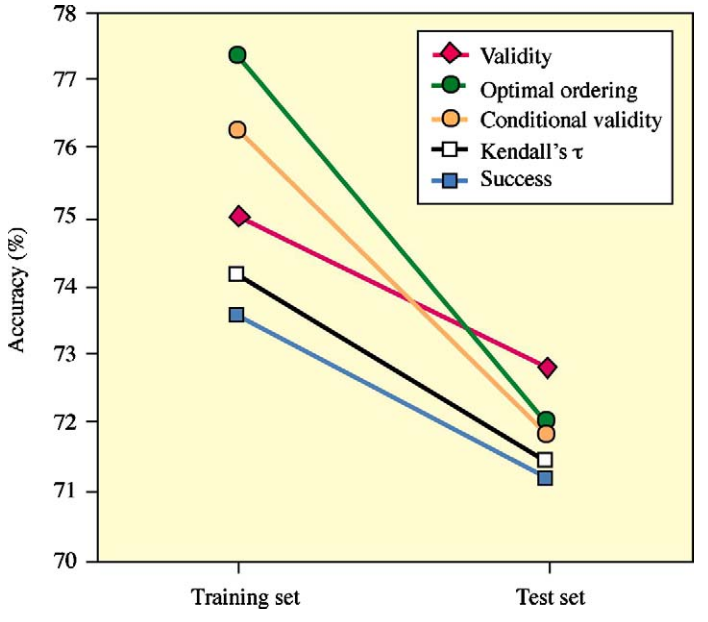

Figure 2. How robust is Take The Best's search rule? Five search rules (for establishing the cue hierarchy of a lexicographic strategy) were tested in a task to infer which of two German cities has a larger population, based on $M = 9$ cues. The reference class consisted of all cities with more than 100,000 inhabitants (83 in all), half of the cities in the training set, and the other half in the test set, yielding about 800 pair comparisons in each set. Performance is averaged across 100 random assignments of cities to training and test sets. Each search rule was used to establish an order in the training set, and a lexicographic strategy with this order was then tested in the test set. Take The Best's search rule orders cues by validity v_i. The optimal ordering is obtained empirically by determining which of all possible orderings of M cues results in the highest accuracy in the training set. The conditional validity of a cue is computed conditionally on the cues that have been looked up before the cue, taking account of the dependencies between cues. Kendall's τ is a rank correlation, which is used here to order cues. Finally, success orders cues according to their probabilistic success. The simple search rule of Take The Best proves to be robust (Martignon and Hoffrage, 2002).

not attempt to compute an optimal stopping point, that is, when the costs of further search exceed its benefits. Rather, the motto of the heuristic is "Take The Best, ignore the rest." The term "one-reason decision making" refers to decision rules that do not weigh and integrate information, but rely on one cue only.

2. Simple Rules for Search

Take The Best orders cues according to their validities, which can be estimated from previous experience (e.g., on a training set). Like the stopping rule, the search rule does not employ optimization calculations either. To order cues according to v_i is fast and

Table 1
Six cues for predicting homelessness in U.S. cities. Cues are ordered by validity, with rent control having the highest (.90) validity (from Tucker, 1987)

	Los Angeles	Chicago	New York	New Orleans
Homeless per million	10,526	6618	5024	2671
Rent control (1 is yes)	1	0	1	0
Vacancy rate (1 is below median)	1	1	1	0
Temperature (1 is above median)	1	0	1	1
Unemployment (1 is above median)	1	1	1	1
Poverty (1 is above median)	1	1	1	1
Public housing (1 is below median)	1	1	0	0

frugal, but not optimal, because this order ignores dependencies between cues. How much more accurate would the optimal order be? Figure 2 shows two unexpected results. First, in a noisy real-world environment, Take The Best actually comes close to the optimal ordering when the task is to fit given data (i.e., training set). Second, and most important, when the task is to predict new data, the simple ordering used by Take The Best is actually more robust and makes more accurate predictions (on the test set) than the ordering that was optimal on the training set. Thus, the simple ordering, which ignores dependencies between cues, turns out to be the better one when generalizing to new objects. The simple search rule of Take The Best strikes a balance between the dangers of overfitting (i.e., extracting too much information from the training set, as optimal ordering and conditional validity do) and underfitting (extracting too little information, which Kendall's τ and success do). In general, a model A overfits the training data if there exists an alternative model B, such that A has higher or equal accuracy than B in the training set, but lower accuracy in the test set.

Minimalist is another heuristic that embodies one-reason decision making. It does not try to order cues by validity, but chooses cues in random order. The only difference from Take The Best is the search rule, which now reads:

STEP 1. Search rule: Pick a cue randomly (without replacement) and look up the cue values of the two objects.

What price does one-reason decision making have to pay for being fast and frugal? How much more accurate are strategies that use all cues and combine them? We first answer these questions for one specific example – homelessness rates – in order to explain the logic of the tests. Thereafter, we report the results of 20 studies, including economic, demographic, environmental, and other prediction tasks.

3. Predicting Homelessness

The task is to predict which of two cities has a higher homelessness rate, using the data on 50 U.S. cities from Tucker (1987). An excerpt from the data, including the values for Los Angeles, Chicago, New York, and New Orleans on six relevant cues, and the homelessness rates, is shown in Table 1. Here, and in subsequent studies reported, there are no unknown objects; thus the recognition heuristic is of no use. One cue (rent control) is binary, and the other five have been dichotomized at the median. For example, cities with rent control more often have a higher homelessness rate than cities without rent control; therefore cities that have rent control are marked with a cue value of "1" for this cue.

In the tests, half of the cities were randomly drawn. From all possible pairs within this training set, the order of cues according to validity v_I was determined. Thereafter, performance was tested on the other half of the cities. Minimalist used the training set only to determine whether a cue is positively or negatively correlated with the criterion (e.g., whether rent control indicates higher or lower homelessness rates). In the test set, it picked the cues in a random order. Two linear models were introduced as competitors: multiple regression and a simple unit-weight linear model (Dawes, 1979). To determine which of two cities has the higher rate, multiple regression estimated the homelessness rates of each city, and the unit-weight model simply added up the number of 1's.

Figure 3, left panel, shows the frugality (average number of cues looked up) and the accuracy of the two fast and frugal heuristics and the two linear models. The two heuristics looked up on average only 2.1 and 2.4 cues, as opposed to 6 cues used by the linear models that have no search and stopping rules. In data fitting (training set), multiple regression fits the data best. The striking result is that Take The Best is more accurate in prediction (test set) than multiple regression and the other competitors. Minimalist also does surprisingly well given the little information it uses.

4. Fast and Frugal Heuristics Versus Linear Models: A Competition

How well do these results generalize to making predictions in other domains? Czerlinski, Gigerenzer, and Goldstein (1999) tested one-reason decision making on 20 prediction problems. These data sets contained real-world structures rather than convenient multivariate normal structures; they ranged from having 11 to 395 objects, and from 3 to 19 cues. The predicted criteria included economic variables, such as selling prices of houses and professors' salaries; demographic variables, such as mortality rates in U.S. cities and population sizes of German cities; environmental variables, such as the amount of rainfall, ozone, and oxidants; health variables, such as obesity at age 18; and sociological variables, such as drop-out rates in Chicago public high schools.

Figure 3, right panel, shows that the counterintuitive results obtained for predicting homelessness held up on average across these 20 different prediction problems. The two fast and frugal heuristics looked up fewer than a third of the cues. Minimalist was most

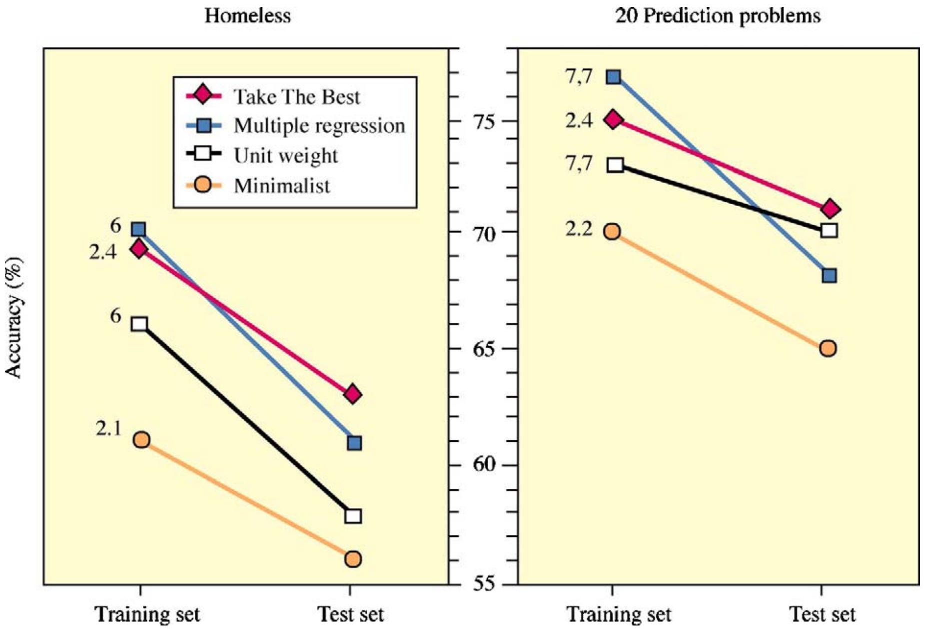

Figure 3. A competition between two heuristics and two linear models. The left panel shows the accuracy and frugality of the four strategies in predicting homelessness in U.S. cities, and the right panel shows the results for 20 real-world problems (Czerlinski, Gigerenzer, and Goldstein, 1999). Accuracy is measured for data fitting (performance in the training set) and prediction (performance in the test set), Take The Best and Minimalist are heuristics that practice one-reason decision making, whereas the unit-weight model and multiple regression use all information available and combine all cues. The numbers next to the graphs denote the average number of cues that have been used by this strategy.

frugal and performed not too far behind the two linear strategies in predictive accuracy (test set). Take The Best was both more frugal and more accurate than the two linear strategies. This result may sound paradoxical because multiple regression processed all the information that Take The Best did and more (we resolve this apparent paradox below).

5. Fast and Frugal Heuristics Versus Bayesian Methods

How does Take The Best compare to Bayesian methods? With large numbers of cues, as with many of the 20 predictive problems studied, Bayes' rule can no longer be used, because it quickly leads to computational explosion. Martignon and Laskey (1999) used two approximations, one simple and one that used days of computing time. The simple Bayesian model was naive Bayes, which assumes that all cues are independent of each other, given the criterion. The computationally expensive model was a Bayesian net-

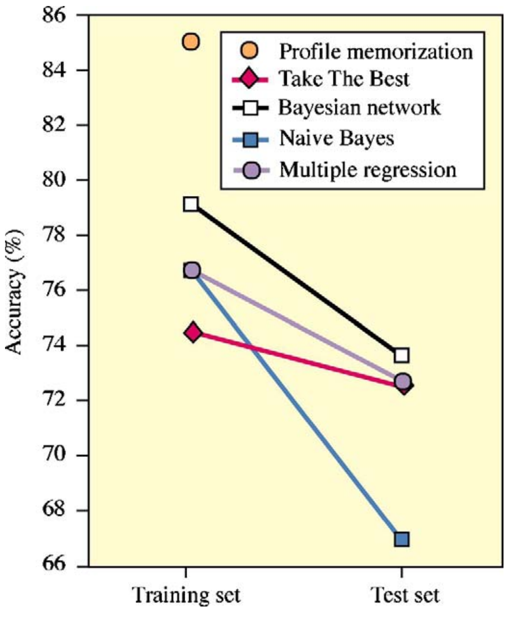

Figure 4. The accuracy of Take The Best compared to a Bayesian network, naive Bayes, and multiple linear regression across 20 predictive problems. The profile memorization method specifies the upper limit of accuracy in the case of fitting given data (Martignon and Laskey, 1999).

work that estimates relevant dependencies between cues from the data. Figure 4 shows that the predictive accuracy of Take The Best came, on average, within three percentage points of the complex Bayesian network, with naive Bayes in-between.

What would be the maximum accuracy a strategy could reach? We can answer this question for fitting known data (i.e., performance in the training set). The optimal Bayesian method for fitting known data – we call it the profile memorization method – memorizes the corresponding criterion value for each cue profile. When comparing two profiles, it chooses the one for which the memorized criterion is larger. If there are several pairs of objects with the same pair of cue profiles, the method determines the proportion of pairs for which the first object scores higher and makes an inference based on whether this proportion is larger than 0.5. For the 20 problems, profile memorization results in a fit of 85% on average. However, this method cannot be used for generalization (test set) because, in new data, unknown profiles may appear.

6. Why is Take The Best so Robust?

The answer lies in its simplicity: Take The Best uses few cues. The first cues tend to be highly valid and, in general, they will remain so across different subsets of the same class of objects. The stability of highly valid cues is a main factor for the robustness of Take The Best, that is, its low danger of overfitting in cross-validation as well as in other forms of incremental learning. In contrast, strategies that use all cues must estimate a number of parameters larger than or equal to the number of cues. Some, like multiple regression, are sensitive to many features of the data, for instance, by taking correlations between cues into account. As a consequence, they suffer from overfitting, especially with small data sets.

The result that simple heuristics can match strategies that use more information is reminiscent of the phenomenon of flat maxima. If many sets of weights, even unit weights, can perform about as well as the optimal set of weights in a linear model, this is called a flat maximum (e.g., Dawes and Corrigan, 1974). The performance of Take The Best indicates that flat maxima can extend beyond weights: Inferences based solely on the best cue can be as accurate as those based on any other weighted linear combination of all cues. The theorems presented below, in particular the theorem on non-compensatory information, identify conditions under which we can predict flat maxima.

7. Ecological Rationality: Which Environmental Structures Can Take The Best Exploit

What are the characteristics of information in real-world environments that make Take The Best a better predictor than other strategies, and where will it fail? To answer these questions, we need to examine properties of information, that is, the information about an environment known to a decision maker. Here we discuss three properties. The first two characterize many real-world situations, at least approximately: When the information structure is non-compensatory, or the available information is scarce, Take The Best is smarter than its competitors. The third property is abundance of information: When information is abundant, a simple unit-weight linear rule will be more accurate.

8. Non-compensatory Information

Among the 20 environments in Figure 3, we found 4 in which the weights for the linear models were non-compensatory (i.e., each weight is larger than the sum of all other weights to come, such as $1/2, 1/4, 1/8, \ldots$). In short, we refer to an environment with such a structure as a non-compensatory environment. Figure 5 shows examples of non-compensatory and compensatory environments. The following theorem states a property of non-compensatory environments and is easily proved (Martignon and Hoffrage, 2002):

THEOREM 1. *Take The Best is equivalent – in performance – to a linear model whose weights form a non-compensatory set* (*and decay in the same order as that of Take The Best*).

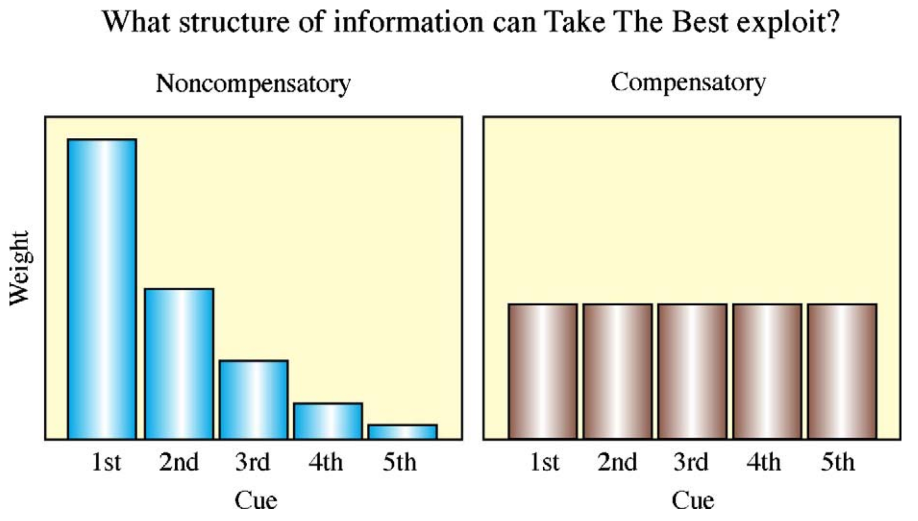

Figure 5. Heuristics can exploit structures of environments. The left side shows an environment which consists of binary cues whose weights are non-compensatory (e.g., $1/2$, $1/4$, $1/8$, and so on). In this environment, no weighted linear model can outperform the faster and more frugal Take The Best. The right side shows a compensatory environment, where linear models will have an advantage (Martignon and Hoffrage, 2002).

Therefore, if an environment consists of cues that are non-compensatory, then no linear model can have higher predictive accuracy than Take The Best.

9. Scarce Information

To illustrate the concept of scarce information, let us recall a fact from information theory: A class of N objects contains $\log_2 N$ bits of information. This means that if we were to encode each object in the class by means of binary cue profiles of the same length, this length should be at least $\log_2 N$ if each object is to have a unique profile. For instance, to encode eight objects, it is sufficient to use three ($\log_2 8 = 3$) binary variables. If there were only two, these eight objects could not be perfectly distinguished, and for some pairs there would be identical cue profiles.

THEOREM 2. *If the number of cues is fewer than* $\log_2 N$, *profile memorization method will never achieve* 100% *correct inferences. Thus, no other strategy will do so either.*

This theorem motivates the following:

DEFINITION 1. A set of M cues provides scarce information for a reference class of N objects if $M \leqslant \log_2 N$.

We can now formulate a theorem that relates the performance of Take The Best to that of a unit-weight linear model in small environments, that is, in environments with fewer than 1000 objects.

THEOREM 3. *In the majority of small environments with scarce information, Take The Best is more accurate than a unit-weight linear model.*

This result was obtained by exhaustive counting. The intuition underlying the theorem is the following: In scarce environments, a unit-weight linear model can take little advantage of its strongest property, namely compensation.

10. Abundant Information

Adding cues to a scarce environment will do little for Take The Best if the best cues in the original environment already have high validity. For a unit-weight linear model, however, adding cues may help because they can compensate for various mistakes this rule would have made if restricted to using only the first cues. In fact, by continually adding cues, we can make a unit-weight linear model achieve perfection. This is true even if all cues are uncertain, that is, if all cues have a validity of less than 1.

THEOREM 4. *If an environment consists of all possible uncertain cues, a unit-weight linear model will discriminate among all objects and make only correct inferences.*

The proof is given in Martignon and Hoffrage (2002). Note that in the context of Theorem 4, we are using the term "cue" to denote a binary-valued function in the reference class. Therefore, the number of different cues in a finite reference class is finite. The theorem can be generalized from the simple linear model with unit weights to linear models that use cue validities as weights.

11. Do People Intuitively Adapt Heuristics to Environmental Structures?

How do people know when to apply which heuristic? Can mere feedback select heuristics? In an experiment by Rieskamp and Otto (2002), participants took the role of bank consultants with the task of evaluating which of two companies applying for a loan was more creditworthy. Six cues such as qualification of employees and profitability were provided for each company. For the first 24 pairs of companies, no feedback was provided as to the correctness of the participants' inferences. Thereafter, feedback was given. For one group of participants, the correct answer was determined in about 90% of the cases by Take The Best, that is, feedback was obtained from the cues in a non-compensatory way. For the second group, the more creditworthy company was determined in about 90% of the cases by a weighted additive rule, that is, the feedback

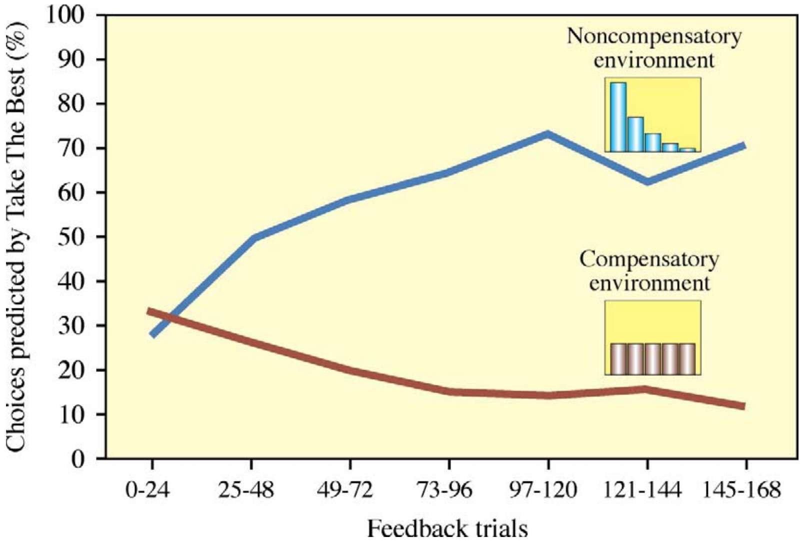

Figure 6. Do people intuitively learn when to use Take The Best? When the participants were in an experimental environment in which feedback was generated in a non-compensatory way, the frequency of choices consistent with Take The Best increased over time; when the feedback was compensatory, this frequency decreased (Rieskamp and Otto, 2002).

was generated in a compensatory way. Did people intuitively adapt their heuristics to the feedback structure of the environments? As can be seen from Figure 6, this was the case: Feedback changed the frequency of responses consistent with Take The Best. Note that in this experiment, participants could acquire information without paying for it. This fosters compensatory strategies, as can be seen from the low initial frequency of around 30% for Take The Best. People learned – without instruction – that different heuristics are successful in different environments.

12. Does the Use of Lexicographic Strategies Depend on Time Pressure?

Empirical evidence for lexicographic strategies (e.g., Payne, Bettman, and Johnson, 1988, 1993; Edland, 1994) and Take The Best (e.g., Bröder, 2000; Newell and Shanks, 2003) has been frequently reported in the literature. Take The Best (but not Minimalist) is a variant of a lexicographic strategy, although it has additional features, including the recognition heuristic as its initial step and an asymmetric stopping rule for unknown values. Rieskamp and Hoffrage (1999) tested how well eight strategies proposed in the literature predicted people's decisions under low and high time pressure. The participants' task was to predict which of four companies had the highest yearly profit. They could look up, sequentially, the information from six cues (e.g., amount of investments, the number of employees, etc.). Two strategies modeled participants' choices

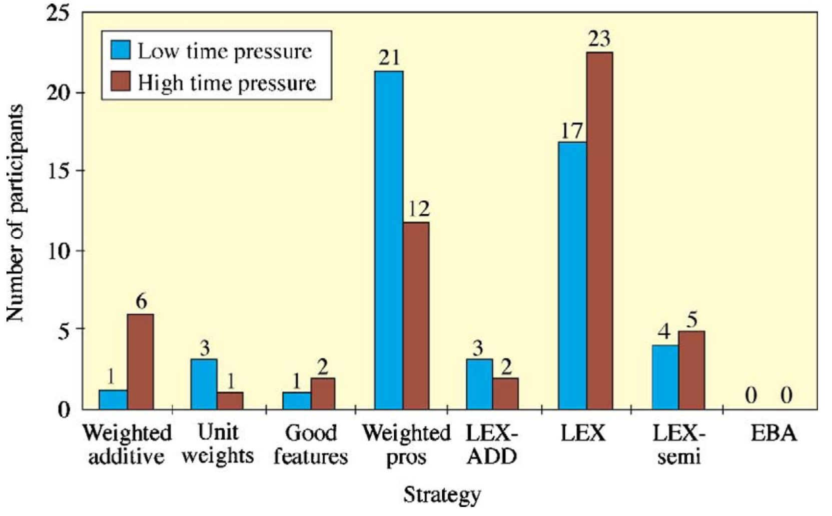

Figure 7. A test of how well eight strategies predict people's choices under low (50-second) and high (20-second) time pressure (Rieskamp and Hoffrage, 1999). Participants had to infer which of four companies had the highest profit. Two strategies, Weighted Pros and LEX, best predicted participants' behavior for low and high time pressure, respectively. LEX, the most simple heuristic among the candidates, is a generalization of Take The Best from binary choices to choices among several alternatives, whereas Weighted Pros is a simple compensatory strategy. The Weighted Additive Model weights cues by their validities and adds all weighted cue values; the Unit-Weight Model attaches the same weight to each cue; Good Features (Alba and Marmorstein, 1987) selects the alternative with the highest number of good features, that is, cue values that exceed a specified threshold. Weighted Pros (Huber, 1979) considers only the highest value on each cue (i.e., ignores all other values) and computes the sum of the validities of these cues for each alternative. LEX-ADD is a two-step strategy: It first uses LEX-Semi to choose two alternatives as favorites, then evaluates them by a unit-weight model, and finally selects the one with the highest score. LEX-Semi (Luce, 1956) works like LEX, except that negligible differences between cue values are disregarded. LEX and LEX-Semi are the only strategies in this set of eight which employ one-reason decision making. Elimination By Aspects (EBA; Tversky, 1972) eliminates all alternatives that do not exceed a specified value on the first cue examined. If more than one alternative remains, another cue is selected.

best: a generalization of Take The Best from binary choices to choices among several alternatives (lexicographic heuristic or LEX) and Weighted Pros (Huber, 1979). Weighted Pros considers only the highest value on each cue (i.e., ignores all other values). Under time pressure, participants' choices conformed better to LEX, which is also the computationally less expensive strategy. Only a low number of participants could be best described by any of the other six models.

13. An Intelligent System Must Ignore Information

This section illustrates both the descriptive validity and the prescriptive power of simple heuristics that employ one-reason decision making. It has long been known that people often base their decisions on only one, or a few, reasons. However, in the behavioral economics literature, ignoring information is all too quickly labeled a judgmental bias, or an act of irrationality. Sometimes this is true, but often it is not, as the performance of Take The Best demonstrates. For instance, in the 1960s, conservatism – that is, the overweighting of base rates in Bayesian problems – was reported, whereas in the 1970s and 1980s, the base rate fallacy – that is, the underweighting or ignoring of base rates – was considered to be an established fact. Both phenomena appeared to contradict each other, and researchers puzzled as to why (Gigerenzer, 2000). It is now easy to see how heuristics such as Take The Best can produce both phenomena. If one of the predictors refers to the base rates of events, then either base rate neglect or conservatism will result, depending on where this base rate cue is in the cue hierarchy. There is no puzzling empirical contradiction, nor is there necessarily irrationality. Rather, by means of precise models of heuristics – unlike merely verbal labels such as availability – we can understand how phenomena and anomalies are created.

References

Alba, J.W., Marmorstein, H. (1987). "The effects of frequency knowledge on consumer decision making". Journal of Consumer Research 14, 14–26.

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, 1332–1346.

Czerlinski, J., Gigerenzer, G., Goldstein, D.G. (1999). "How good are simple heuristics?" In: Gigerenzer, G., Todd, P.M., the ABC Research Group (Eds.), Simple Heuristics that Make Us Smart. Oxford University Press, New York, pp. 97–118.

Dawes, R.M. (1979). "The robust beauty of improper linear models in decision making". American Psychologist 34, 571–582.

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

Edland, A. (1994). "Time pressure and the application of decision rules: Choices and judgments among multiattribute alternatives". Scandinavian Journal of Psychology 35, 281–291.

Gigerenzer, G. (2000). "Adaptive Thinking: Rationality in the Real World". Oxford University Press, New York.

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

Gigerenzer, G., Todd, P.M., the ABC Research Group (1999). "Simple Heuristics that Make Us Smart". Oxford University Press, New York.

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

Huber, O. (1979). "Nontransitive multidimensional preferences: Theoretical analysis of a model". Theory and Decision 10, 147–165.

Luce, R.D. (1956). "Semiorders and a theory of utility discrimination". Econometrica 24, 178–191.

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

Martignon, L., Laskey, K. (1999). "Bayesian benchmarks for fast and frugal heuristics". In: Gigerenzer, G., Todd, P.M., the ABC Research Group (Eds.), Simple Heuristics that Make Us Smart. Oxford University Press, New York, pp. 169–188.

Newell, B., Shanks, D. (2003). "Take the best or look at the rest? Factors influencing 'one-reason' decision-making". Journal of Experimental Psychology: Learning, Memory, and Cognition 29, 53–65.

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, 534–552.

Payne, J.W., Bettman, J.R., Johnson, E.J. (1993). "The Adaptive Decision Maker". Cambridge University Press, New York.

Rieskamp, J., Hoffrage, U. (1999). "When do people use simple heuristics, and how can we tell?" In: Gigerenzer, G., Todd, P.M., the ABC Research Group (Eds.), Simple Heuristics that Make Us Smart. Oxford University Press, New York, pp. 141–167.

Rieskamp, J., Otto, P. (2002). "Adaptive strategy selection in decision making: A learning perspective". Manuscript in preparation.

Tucker, W. (1987). "Where do the homeless come from?" National Review 25, 34–44.

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