

## THEORETICAL NOTES

# The Robust Beauty of Ordinary Information

Konstantinos V. Katsikopoulos

Max Planck Institute for Human Development, Berlin, Germany,  
and Massachusetts Institute of Technology

Lael J. Schooler

Max Planck Institute for Human Development,  
Berlin, Germany

Ralph Hertwig

University of Basel

Heuristics embodying limited information search and noncompensatory processing of information can yield robust performance relative to computationally more complex models. One criticism raised against heuristics is the argument that complexity is hidden in the calculation of the cue order used to make predictions. We discuss ways to order cues that do not entail individual learning. Then we propose and test the thesis that when orders are learned individually, people's necessarily limited knowledge will curtail computational complexity while also achieving robustness. Using computer simulations, we compare the performance of the take-the-best heuristic—with dichotomized or undichotomized cues—to benchmarks such as the naïve Bayes algorithm across 19 environments. Even with minute sizes of training sets, take-the-best using undichotomized cues excels. For 10 environments, we probe people's intuitions about the direction of the correlation between cues and criterion. On the basis of these intuitions, in most of the environments take-the-best achieves the level of performance that would be expected from learning cue orders from 50% of the objects in the environments. Thus, ordinary information about cues—either gleaned from small training sets or intuited—can support robust performance without requiring Herculean computations.

*Keywords:* inductive inference, heuristics, take-the-best, robustness, bootstrapping

A psychological regularity touted as a potentially universal law of cognition is the effort–accuracy tradeoff: Only if people invest more cognitive effort do they stand to achieve more accuracy in their choices and judgments. More effort can take the form of searching for information exhaustively, spending plenty of time on the problem, or performing complex computations. Yet, this law may fall short of being universal. Inspired by the ideas of Herbert Simon, those in the research program on fast and frugal heuristics have developed computational models of heuristics as one precise interpretation of Simon's (1956) notion of *bounded rationality*. Although computational models of heuristics were proposed before (e.g., Payne, Bettman, & Johnson, 1993; Tversky, 1972), they were built on the premise that people rely on heuristics because they lack the cognitive

capacity to perform rational calculations or are willing to sacrifice accuracy by expending less effort. This view has been challenged by demonstrations that processes embodying bounded rationality, such as limited information search and noncompensatory processing, can lead to more accurate inferences than can be achieved by models based on more information and complex computations (Gigerenzer, Todd, & The ABC Research Group, 1999).

This challenge to the effort–accuracy tradeoff, however, has not remained undisputed. One important objection suggests that the heuristics' success rides on complex computations. For example, many heuristics, such as take-the-best (Gigerenzer et al., 1999), do not use all available *cues* (i.e., variables that correlate, albeit imperfectly, with the criterion variable in a decision problem; see Brunswik, 1955) but instead order them, look them up one by one, and stop searching as soon as a discriminating cue is encountered. Juslin and Persson (2002, p. 575) stressed that much effort is needed to order cues. For example, in Gigerenzer et al.'s (1999) city-population-comparison task (e.g., which of the following two cities has more residents, Heidelberg or Bonn?), the computation of cue order involving 83 German cities and nine cues requires 30,627 comparisons (for details, see Juslin & Persson, 2002). Others have raised similar concerns (Dougherty, Franco-Watkins, & Thomas, 2008; Newell, 2005; Rakow, Hinest, Jackson, & Palmer, 2004).

Do simple heuristics really freeloader on the effort hidden in the computation of cue orders? A first response to the "hidden effort" criticism is that a person can conveniently piggyback on the experiences of others, past or present. Evolution, culture, and other individuals can learn cue orders (for details, see Gigerenzer, Hoffrage, &

---

This article was published Online First September 6, 2010.

Konstantinos V. Katsikopoulos, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany, and Department of Mechanical Engineering and Engineering Systems Division, Massachusetts Institute of Technology; Lael J. Schooler, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development; Ralph Hertwig, Faculty of Psychology, University of Basel, Basel, Switzerland.

We thank Anna Cullen for her help with data entry and processing, Carola Fenselow for programming the simulations, members of the ABC Research Group for their comments and helpful discussions, and Laura Wiles for editing the article.

Correspondence concerning this article should be addressed to Konstantinos V. Katsikopoulos, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: katsikop@mpib-berlin.mpg.de

Goldstein, 2008). Of course, despite evolution, culture, and the vicarious experience of others, sometimes a person finds herself with nobody to consult or imitate. This, however, does not dictate that the heuristics this person would use require massive effort for computing cue orders, for two reasons. First, some heuristics do without cue order. For example, the recognition heuristic (Goldstein & Gigerenzer, 2002) and the fluency heuristic (Hertwig, Herzog, Schooler, & Reimer, 2008; Schooler & Hertwig, 2005) need only one cue, benefiting from automatic, albeit potentially computationally complex, memory processes. Other heuristics, such as the priority heuristic, rely on risk preferences to order cues (Brandstätter, Gigerenzer, & Hertwig, 2006), and some heuristics bypass cues by relying on the frugal retrieval of exemplars (Juslin & Persson, 2002). A second reason why heuristics do not demand massive investment in the calculation of cue orders is that minute samples of knowledge can be used, which drastically restricts the computations required. Of course, to paraphrase Alexander Pope, a little cue knowledge can be a dangerous thing. It makes computations workable, but it may lead to ineffective cue orders. Whether severely limited knowledge is a curse or a blessing is the subject of our investigations.

We will test how well simple heuristics perform with ordinary cue information. In the context of this article, ordinary information refers to a person's severely limited knowledge or a person's knowledge of the sign of the correlation between a cue and a criterion, called the cue's direction. Cue directions often follow people's basic conceptions about how the world works: For example, cities with a high unemployment rate (cue) tend to have a high homelessness rate (criterion).

We present two investigations that combine computer simulations and human data. The first analyzes the performance of heuristics that use minute samples. The second analyzes the performance of heuristics that rely only on people's intuitions about cue directions. We next describe the heuristics.

### Fast and Frugal Heuristics

We focus on the inference models studied by Gigerenzer et al. (1999): three heuristics (tallying, take-the-best, and minimalist) and two benchmarks (linear regression and naïve Bayes). The task is a paired comparison, in which the goal is to predict which of two objects has the higher value on a criterion. For example, it can be asked which of two cities, Chicago or Los Angeles, has a higher homelessness rate. Lacking direct knowledge of these values, most people have to use cues. For example, the cues associated with a city's homelessness rate could be the average temperature (a continuous cue) and the presence of a university (a binary cue).

These models differ in how they process cues. Linear regression weighs and sums cues, predicting that the object with the higher sum has the higher criterion. Naïve Bayes selects the object with the higher probability of having the higher criterion value, given the objects' values on the cues. In contrast, the three heuristics represent simplifications of weighted linear models. The first one, *tallying* (Dawes, 1979), dispenses with weighting and simply adds binary cue values. The *take-the-best* heuristic dispenses with adding and weighs cues simply by ordering them by a measure of cue goodness called *validity* and looking them up one at a time. As soon as a discriminating cue is found, search is stopped and a decision is made on the basis of that cue alone. *Minimalist* operates like take-the-best except that it randomly orders cues.

In sum, relative to the heuristics, the benchmarks use more complex computations and sophisticated information. Multiple regression and naïve Bayes use the precise *values* of regression weights and cue validities, whereas take-the-best uses the *order* of cue validities. Minimalist and tallying do not require cue validities or orders but only cue *directions*.

Czerlinski, Gigerenzer, and Goldstein (1999) applied these inference models to 20 data sets from domains such as biology, economics, and sociology. Each data set was split in half, and the parameters of each model (weights, validities, directions) were estimated on half of the data, the *training set*. The parameter estimates were used to predict, for each model, the paired comparisons on the other half of the data, the *test set*. This process was repeated 1,000 times. In the test set, the predictive accuracy of the models was as follows: minimalist 65%, regression with dichotomized cues 68%, tallying 69%, take-the-best with dichotomized cues 71%, naïve Bayes 73%, regression with undichotomized cues 76%, and take-the-best with undichotomized cues 76%.

Using a city-population-comparison task, Juslin and Persson (2002) and Chater, Oaksford, Nakisa, and Redington (2003) provided some evidence that heuristics have high predictive accuracy when the training set includes fewer than 50% of all objects. For example, Chater et al. tested take-the-best against a three-layer feed-forward connectionist network, two exemplar-based models, and a decision-tree-induction algorithm. With training sets of up to 40% of all objects, take-the-best outperformed or matched its competitors but was outperformed with larger training sets. Our first goal is to find out how robust these initial results are across many environments and with minute training set sizes with as few as two objects.

### Individual Learning From Minute Samples: How Good Could It Be?

Table 1 lists the nine models included in the first simulation: two versions of multiple regression, one in which continuous cues were *dichotomized* ( $MR_D$ ) and one in which they were not ( $MR_U$ ); tallying (TAL) and minimalist (MIN), which both use dichotomized cues, and three versions of take-the-best, one that uses undichotomized cues ( $TTB_U$ ) and two that use dichotomized cues. The latter two versions differ in their computation of validity, which can be done by a *frequentist* ( $TTB_F$ ) or a *Bayesian* ( $TTB_B$ ) approach (in the undichotomized take-the-best version we used frequentist validity). We also included two versions of naïve Bayes: one with frequentist ( $NB_F$ ) and one with Bayesian ( $NB_B$ ) validity. The equation for the frequentist cue validity is  $v = R/(R + W)$ , where  $R$  and  $W$  are the number of correct and incorrect comparisons, respectively, based on the cue. The frequentist approach may weight the first observations too much, compromising performance on minute samples.<sup>1</sup> The Bayesian approach overcomes this problem (Lee & Cummins, 2004) by calculating validity as  $v = (R + 1)/(R + W + 2)$ .

<sup>1</sup> This is a problem with some inference models, such as, for example, naïve Bayes. Naïve Bayes selects the object with the larger product  $\prod_i v_i(1 - v_j)$ , where  $v_i$  is the validity of any cue  $c_i$  that has a higher value on the object, and  $v_j$  is the validity of any cue  $c_j$  that has a higher value on the other object. If validities equal 0 or 1, the product will equal 0 for both objects (unless all cues with validity 1 point to one object and all cues with validity 0 point to the other), and naïve Bayes will have to guess. In line with this point, naïve Bayes with frequentist validity scored 50% for a training set of two objects (see Table 1).

Table 1  
*Summary of the Fast and Frugal Heuristics and Benchmark Models Tested in the First Investigation*

Model	Decision rule	Parameters
Multiple linear regression with dichotomized cues (MR <sub>D</sub> )	Select object with higher weighted sum of cue values	Regression weights
Multiple linear regression with undichotomized cues (MR <sub>U</sub> )	Select object with higher weighted sum of cue values	Regression weights
Naïve Bayes with dichotomized cues and frequentist validity (NB <sub>F</sub> )	Select object with higher probability of having higher criterion value	Cue validities
Naïve Bayes with dichotomized cues and Bayesian validity (NB <sub>B</sub> )	Select object with higher probability of having higher criterion value	Cue validities
Tallying with dichotomized cues (TAL)	Select object with higher sum of cue values	Cue directions
Minimalist with dichotomized cues (MIN)	Select object with higher value on random cue that discriminates	Cue directions
Take-the-best with dichotomized cues and frequentist validity (TTB <sub>F</sub> )	Select object with higher value on most valid cue that discriminates	Order of cue validities
Take-the-best with dichotomized cues and Bayesian validity (TTB <sub>B</sub> )	Select object with higher value on most valid cue that discriminates	Order of cue validities
Take-the-best with undichotomized cues and frequentist validity (TTB <sub>U</sub> )	Select object with higher value on most valid cue that discriminates	Order of cue validities

**Design**

The predictive accuracy of the inference models was tested in 19 of the 20 data sets used by Czerlinski et al. (1999; we omitted a data set with fewer than 20 objects). To compare our results to previous work, we investigated a training set that included 50% of all objects. We also investigated training sets ranging in size from two to 10 objects. Averaged across the data sets, training sets of two and 10 objects equal 3% and 15% of all objects, respectively.

The models, with parameters (weights, validities, directions) estimated from all objects in the training set, were applied to the test set. When a model required binary cues, these were dichotomized cues using the sample median. The simulation was repeated 5,000 times for each model, environment, and training set size.

**Results**

Figure 1 plots mean predictive accuracy, defined as a model’s proportion of correct inferences in the test set as a function of the training set size. TTB<sub>U</sub> performs best in all cases except when the training set has two objects. In this case, TTB<sub>U</sub> is slightly behind TAL. For training sets with five to 10 objects, TTB<sub>U</sub> surpasses the second-best model, NB<sub>B</sub>, by at least 5%. With seven objects, TTB<sub>U</sub> reaches 74%—near its maximum performance of 76%, achieved with a training set of 50% of all objects.

Bayesian validities guard against the extreme estimates that the frequentist approach may produce for very small training sets. Unlike take-the-best, naïve Bayes depends on point estimates of validity, so Bayesian validities could be important for its performance. With a training set of two objects, NB<sub>B</sub> performs substantially better than does NB<sub>F</sub> (58% vs. 50%; see also Footnote 1). This advantage disappears with 10 objects or more. The other benchmark, linear regression, does not fare well for very small training set sizes. This is not a new finding (Dawes, 1979; Hogarth & Karelaia, 2005). MR<sub>D</sub> does not outperform the MIN heuristic for any size of the training set. MR<sub>U</sub> surpasses MIN only when there are six objects in the training set, and TAL only for nine objects in the training set.

In sum, we found that the models are differentially influenced by the information that minute samples afford. The predictive

accuracy of the two benchmarks, linear regression and naïve Bayes, is compromised, presumably because training sets with 10 or fewer objects provide unreliable point estimates of regression weights and cue validities. Still, the use of Bayesian estimates provides a decent remedy for naïve Bayes. The heuristics, in contrast, seem capable of making do with very limited information: TAL is the most robust model for training sets with two objects, whereas TTB<sub>U</sub> is the most robust model for sets including between three and 10 objects, with a difference in accuracy of more than 5 percentage points on average. In addition, with as few as seven objects TTB<sub>U</sub> reaches a performance level close to the maximum achieved by any model with a 50% training set. Hogarth and Karelaia (2007) conjectured that heuristics such as TTB<sub>U</sub> “would be less sensitive to sampling errors” (p. 751). This is exactly what we found. Why is it so?

**The Robust Beauty of Cue Directions**

One possible reason for the robustness of heuristics is that even minute samples identify a few good cues. For example, could it be that at least one of the, say, three most valid cues is ordered correctly? No. For training sets with two to eight objects, it is more likely that the three most valid cues are ranked incorrectly than that at least one of them is ranked correctly. Interestingly, getting the validity order right does not really matter much, as shown in Figure 2.

In the upper panel of Figure 2 we plot, for all training set sizes, the correlation between the predictive accuracy of take-the-best (and MIN) and how closely the validity order estimated from the training set corresponds to the validity order in the population. For each one of the 5,000 repetitions of the simulation and for each training set size, we measured the degree of correspondence between the two orders by a rank correlation. We separately considered the three most valid cues and all cues. None of the correlations between predictive accuracy and degree of correspondence reach a value higher than .2 for any training set size.

If cue order does not affect predictive accuracy much, then what does? Cue directions do. In the lower panel of Figure 2, we graph, for all training set sizes, the correlation between the predictive

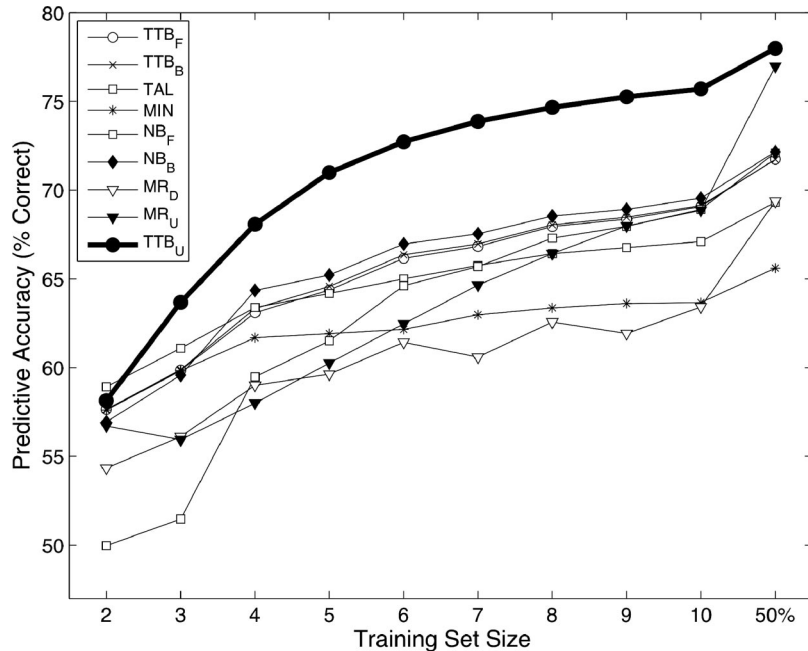


Figure 1. Mean predictive accuracy (across 19 environments) of fast and frugal heuristics and benchmark models as a function of the size of the training set (2–10 objects and 50% of the objects in the population). TTB<sub>F</sub> = take-the-best with dichotomized cues and frequentist validity; TTB<sub>B</sub> = take-the-best with dichotomized cues and Bayesian validity; TAL = tallying with dichotomized cues; MIN = minimalist with dichotomized cues; NB<sub>F</sub> = naïve Bayes with dichotomized cues and frequentist validity; NB<sub>B</sub> = naïve Bayes with dichotomized cues and Bayesian validity; MR<sub>D</sub> = multiple linear regression with dichotomized cues; MR<sub>U</sub> = multiple linear regression with undichotomized cues; TTB<sub>U</sub> = take-the-best with undichotomized cues and frequentist validity.

accuracy of take-the-best (and MIN) and how closely the cue directions estimated from the training set correspond to the cue directions in the population. We measured the degree of correspondence between these two sets of cue directions by the number of cue directions in the population that are correctly estimated from the training set, for each one of the 5,000 repetitions of the simulation and for each training set size. As before, we considered the three most valid and all cues. The correlations start very high, between .6 and .7 for a training set with two objects. As the size of the training set increases, correlations decrease, possibly because there is increasingly less variability in the number of correctly estimated cue directions.

For all training set sizes, we calculated, using all data sets and all simulation repetitions per data set, the probability that the direction of a cue was estimated correctly. To avoid overestimating this probability, we used all repetitions, not only those in which a cue discriminated and its direction could be judged. Even so, the probability is high. By sampling two objects, cue direction is estimated correctly with a probability of about .6, surpasses .7 for six objects, and reaches .8 for 10 objects.

Although the performance of both take-the-best and MIN correlates highly with the correct estimation of cue directions, MIN's poor performance in Figure 1 suggests that cue order also matters. It is better to look up valid cues first, as take-the-best does, rather than look them up randomly, as MIN does. For example, in training sets with two objects there is a .17 probability that the cue looked up first is the most valid (the probability that this would

happen by chance equals .13); and this probability increases to .3 for a training set with four objects, .51 for a set with eight objects, and .56 for a set with 10 objects.

In sum, minute samples do not veridically represent the population validity order. Yet, they afford an ordinary but informative piece of information—the direction of a cue. Moreover, even minute samples offer a slightly better-than-chance probability of looking up good cues first. This, together with good cue directions, explains the impressive robustness of take-the-best.

### Cue Directions: Individual and Collective Intuitions

Our first investigation established that fast and frugal heuristics do not necessarily need good cue orders, relying instead on good cue directions to make accurate inferences. In our second investigation, we ask: How good or bad are people's intuitions about cue directions? The father of the TAL heuristic, Robyn Dawes (1979), wrote that "people are good at picking out the right predictor variables and at coding them in such a way that they have a conditionally monotone relationship with the criterion" (p. 573). But it is known that individuals often develop false beliefs about how variables are associated. For example, Chapman and Chapman (1969) found that students perceived positive correlations between people's interpretations of ink blots and homosexuality, although in reality the correlation in the experimental material was negative.

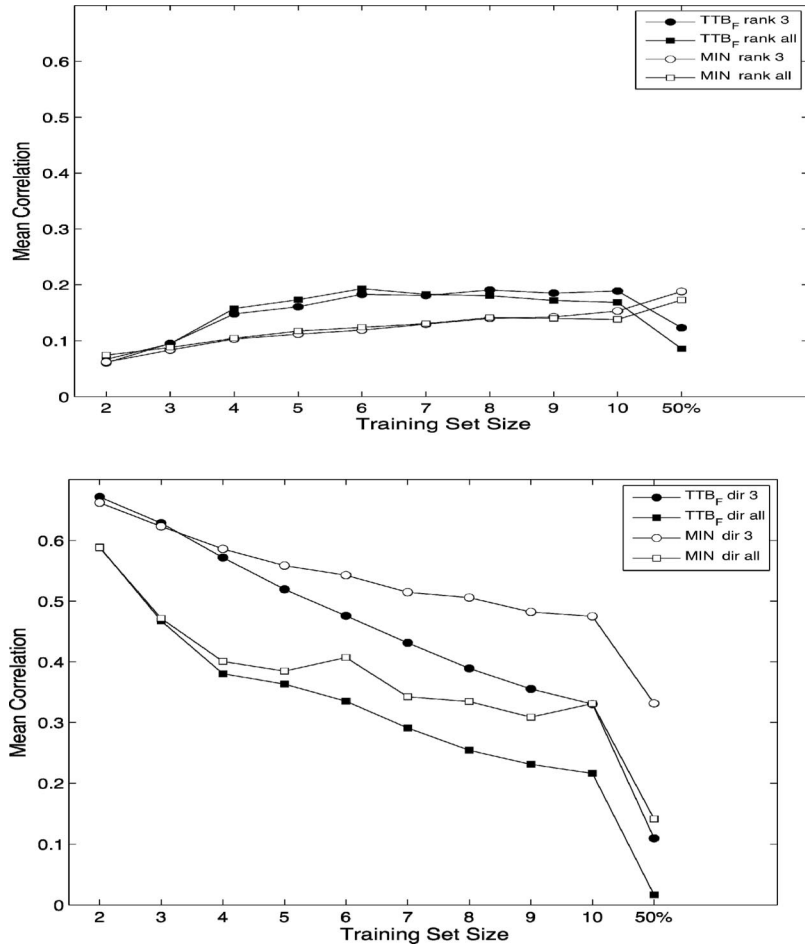


Figure 2. Upper panel: Mean correlations (across 19 environments), as a function of the size of the training set (2–10 objects and 50% of the objects in the population), between the predictive accuracy of take-the-best with dichotomized cues and frequentist validity (TTB<sub>F</sub>) and minimalist with dichotomized cues (MIN) and how closely the cue order estimated from the training set approximates the cue order in the population for the three most valid cues and for all cues. Lower panel: Mean correlations (across 19 environments), as a function of the size of the training set (2–10 objects and 50% of the objects in the population), between the predictive accuracy of TTB<sub>F</sub> and MIN and the number of cue directions in the population that are correctly estimated from the training set for the three most valid cues and for all cues.

We empirically probed people’s intuitions about the directions of cues on 10 of the data sets used before. To see whether these intuitions can lead heuristics to perform well, we also tested the performance of *bootstraps* of heuristics. That is, we used the cue directions judged by participants as input to take-the-best and TAL. Bootstrapping may refer to using the judgments of a person as input to the model of the same person. We use the term more broadly, including the use of an aggregate judgment of a group (defined later) as input.<sup>2</sup>

**Method**

Of the data sets used previously, we eliminated some that are obscure (e.g., predicting rainfall from cloud seeding) or strongly overlap with others (e.g., we selected predicting obesity instead of percentage of body fat). Using the same constraints, we reduced

the cues to an average of 5.1 cues per data set, varying from three to eight, ending up with a total of 51 continuous and binary cues. Table 2 lists the remaining environments and cues. For two environments, participants were likely to have some direct experience (German city populations and body fat in men). It is unlikely, however, that our German participants had experience with the other environments, such as the housing market in Erie, Pennsylvania.

Rather than asking our participants, 50 students from three universities in Berlin, to state the sign of a correlation between cue and criterion, we asked them whether an increase in the value of a specific cue would be associated with an increase or a decrease in

<sup>2</sup> We are not using the term *bootstrap* to refer to the statistical procedure based on resampling.

Table 2  
*10 Environments and 51 Cues Used to Test People's Intuitions About Cue Directions in the Second Investigation*

Environment	Cues
Predict dropout rate of the 57 Chicago public high schools	Grade, percentage of non-White students, percentage of low-income students
Predict the rate of homelessness in 50 U.S. cities	Percentage of public housing, unemployment rate, average temperature, vacancy rate, rent control, percentage of inhabitants below poverty line
Predict the mortality rate in 20 U.S. cities	Percentage of non-White people in urbanized areas, pollution level, average January temperature
Predict populations of the 83 German cities with at least 100,000 inhabitants	City in former East Germany, state capital, exposition site, university, in industrial belt, intercity train connection, license plate, soccer team in premier league
Predict the selling prices of 22 houses in Erie, Pennsylvania	Age of house, number of bathrooms, lot size, living space, property tax, number of bedrooms, garage
Predict the accident rate per million vehicle miles for 37 segments of a highway	Number of lanes, speed limit, percentage of trucks, shoulder width, segment length, lane width, intersections, volume, average traffic count
Predict the average motor fuel consumption per person for each of the 48 contiguous U.S. states	Population, miles per highway, motor fuel tax, percentage of population with license, number of licensed drivers
Predict percentage of body fat for 218 men	Weight, leg circumference, height
Predict the average amount of sleep time for 35 species of mammals	Maximum life span, brain weight, body weight, gestation index
Predict the number of species on 26 Galapagos islands	Distance to adjacent island, distance to coast, area of adjacent island, elevation

the criterion value. We also asked them to order cues by how useful they thought the cues would be. The order in which environments were presented and the order in which cues were presented within environments were varied across participants. For some cues, some participants did not answer, but we collected more than 45 judgments for each cue.

### How Good Are People's Intuitions About Cue Directions?

Across all 10 environments, people judged cue directions correctly in 67% of all cases. The directions of about half of the cues (27 out of 51) were judged correctly by more than 85% of participants. Directions of about three fourths of the cues (38 out of 51) were judged correctly by more than 50% of the participants. Cues judged incorrectly by more than 50% of the participants occur mostly in three environments: mammals' sleep (four cues), highway accidents (three cues), and fuel consumption (three cues). People's incorrect intuitions are not surprising. For example, the number of drivers with a license in a state is negatively correlated with the average fuel consumption in the state. Other cue directions may be justifiable in retrospect but are still surprising, such

as, for example, the number of car accidents on a road segment being negatively correlated with the speed limit. Finally, some cue directions simply require expert knowledge. For instance, it is not clear, to dilettantes in biology such as the authors, why the gestation index is negatively correlated with how much a mammal sleeps.

### How Well Do Heuristics Perform When Fed With People's Cue Directions?

To answer this question, we constructed and tested bootstrap models of take-the-best and a benchmark that we call *calibrated take-the-best* model. The predictive accuracy of the calibrated take-the-best is the predictive accuracy of  $TTB_F$  as analyzed in the first simulation. Across the 10 environments and 51 cues listed in Table 2,  $TTB_F$  was calibrated on a training set with 50% of all objects in each environment. The predictive accuracy of the first bootstrap, which we call *individual take-the-best*, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions and cue orders probed from one of the 50 participants. The second bootstrap, *social take-the-best*, used the cue directions of the majority of participants (García-Retamero, Takezawa, & Gigerenzer, in press). To determine cue order, social take-the-best exploits the strength of the majority about cue directions. For example, if 60% of participants agreed on the direction of one cue, whereas 70% of participants agreed on the direction of another cue, the second cue would be ranked higher.<sup>3</sup>

We also conducted the corresponding analysis for TAL, testing two bootstrap models and a benchmark, *calibrated TAL*. The predictive accuracy of the first bootstrap, *individual TAL*, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions of one of the 50 participants. The second, *social TAL*, takes the direction of a cue to be the direction judged by the majority of the participants (García-Retamero et al., in press).

Table 3 shows the performance of the three take-the-best and three TAL models. Across all 10 environments, the bootstrap models lag from 7 to 10 percentage points behind the calibrated model. Table 2 also shows how accuracy depends on the environment. We distinguish post hoc between *counterintuitive* environments, in which cue directions proved difficult for people to judge (the mammals' sleep, highway accidents, and fuel consumption environments), and *intuitive* environments, in which cue directions were easier to judge (the other seven environments). In counterintuitive environments, for both take-the-best and TAL, the calibrated model does much better than the bootstrap models. In intuitive environments, however, the bootstrap models, especially the social bootstraps, do well and catch up or even outperform the

<sup>3</sup> We also implemented a version of social take-the-best that uses people's intuitions about not only cue directions but also cue orders. It uses Borda counts, the sum of the ranks that a cue was given by all participants, as a proxy for cue ranks (García-Retamero et al., in press). This model reached essentially the same predictive accuracy as did social take-the-best with cue directions and the majority rule.

Table 3  
*Predictive Accuracy (%) of the Models in the Simulation in the Second Investigation*

Model	Predictive accuracy		
	All 10 environments	3 Counterintuitive environments	7 Intuitive environments
Take-the-best			
Calibrated	68	68	68
Individual	58	42	65
Social	60	42	68
Tallying			
Calibrated	67	69	66
Individual	59	46	64
Social	60	41	68

*Note.* The predictive accuracy of a calibrated model and two bootstrap (called individual and social) models, separately for take-the-best and tallying. The calibrated models use all cue-related information in a training set with 50% of all objects in the population, whereas the bootstrap models use cue direction information provided by 50 people (see text for details). Results are provided across all environments and separately for those environments where cue directions were found to be difficult for people to judge (counterintuitive environments) and easy for people to judge (intuitive environments).

calibrated model with perfect knowledge about half of the environment.<sup>4</sup>

## General Discussion

In what follows, we first discuss why getting cue directions right appears to be so crucial for the performance of fast and frugal heuristics, and then we turn to the role of causal theories underlying intuitions about cue directions.

### The Robust Beauty of Ordinary Information

The title of our article alludes to research by Dawes, Hastie, and Kameda on the “robust beauty” of TAL and its group analogue, the majority rule (Dawes, 1979; Hastie & Kameda, 2005). Consistent with the results of their work, the success of TAL in our simulations shows that aggregating cue directions is a robust strategy. What is puzzling about the remarkable performance of take-the-best with undichotomized cues is that it dispenses with aggregation. It relies on just the direction of literally a single cue, the one with the highest validity estimate on the basis of the training set. In the case of continuous cues, this cue will almost always discriminate between two objects. A full solution to this puzzle is still lacking, but we can gain some insight by using two approaches taken in the past to explain the strong performance of heuristics.

First, heuristics achieve high accuracy if the cue structure is noncompensatory. There are at least two meanings of a noncompensatory cue structure, which have been studied analytically: Katsikopoulos and Martignon (2006) showed that, essentially, if the cue with the highest validity has a much higher validity than do all other cues, then take-the-best with binary cues has maximum accuracy among all possible inference models. And Hogarth and Karelaia (2005) showed that, essentially, if the variance in criterion values accounted for by the cue that take-the-best looks up first is higher than the variance in criterion values accounted for by

using all cues in a linear regression (adjusting for the total number of cues), then take-the-best with undichotomized cues is more accurate than regression is.

We offer a speculation for the performance of take-the-best with undichotomized cues rooted in how dichotomization affects validities. One might expect the validity of a continuous cue, under conditions to be specified, to decrease when the cue is dichotomized. The values of two objects on a continuous cue can (a) both lie on the same side of the threshold used to dichotomize the cue or (b) lie on the opposite sides of this threshold. In (b), the validity of the cue is unaffected by dichotomization. In (a), a dichotomized cue does not discriminate, whereas an undichotomized cue does. For environments where the validity of a continuous cue in (a) is higher than its validity in (b), dichotomization lowers cue validity. Irwin and McClelland (2003) made similar observations about the statistical power lost from dichotomizing continuous cues.

In a second approach to explaining the strong performance of take-the-best, Gigerenzer and Brighton (2009) used the fact that the predictive accuracy of any inference model is the sum of its predictions’ bias and variance. They conjectured that heuristics achieve high predictive accuracy because their comparatively low variance compensates for their comparatively high bias. Extrapolating from this account, we suspect that take-the-best with undichotomized cues will have lower variance in its predictions than will take-the-best with dichotomized cues. When cues are continuous, take-the-best determines the relative ranks of two objects, according to the criterion, on the basis of the first cue only. When cues are dichotomized, however, the ranking of objects can, in the most extreme case, depend on the complete order of cues. That is, with three continuous cues there would be, in principle, three rankings of the objects, whereas with three dichotomized cues there would be, in principle, six (3!). This difference could decrease the variance in take-the-best’s predictions with continuous cues, relative to dichotomous cues.

### Cue Directions: Occam’s Razor and Causal Theories

Entities must not be multiplied beyond necessity. If so, what is gained by explaining a person’s judgment for one kind of binary choice, say, which of two mountains in the Swiss Alps is higher, on the basis of another binary variable, intuitions about cue direction? Clearly the most direct solution to the inference problem is to retrieve the mountains’ height directly from memory (see the concept of “local mental models” in Gigerenzer, Hoffrage, & Kleinbölting, 1991). However, there will be situations in which there is no direct knowledge about the criterion. Then, a person may vicariously substitute direct knowledge with probabilistic cues. Our analyses have shed light on the key role of cue directions in this process of substitution.

<sup>4</sup> The results on social TAL can be analytically justified: TAL with correct cue directions can be viewed as a majority rule with jurors with at-least-chance accuracy (Katsikopoulos & Martignon, 2006). The Condorcet jury theorem (Condorcet, 1785) delineates conditions under which the majority rule (TAL) can have high or low accuracy. It says that adding jurors (cues) with at-least-chance accuracy leads to a drastic increase in the accuracy of the majority rule; conversely, adding jurors with below-chance accuracy leads to a drastic decrease in the accuracy of the majority rule. This is what we found.

We speculate that people could arrive at intuitions about cue directions on the basis of their causal knowledge about how the world works. According to Waldmann, Hagmayer, and Blaisdell (2006), people have a “natural capacity to form causal representations” (p. 310), based, in part, on regularities such as “causes typically temporally precede their effects” (p. 308). García-Retamero, Hoffrage, and Dieckmann (2007) have argued that causal representations help people focus on causal associations between cues and criteria, which are likely to be robust. They demonstrated that people were faster to learn causal associations, relative to equally valid associations that were not presented to participants as causal, and that they were more likely to search for cues that were causally related to criteria.

Tentative evidence that causal theories shaped our participants’ intuitions about cue directions can be found in one of the common errors in judged cue directions. For inferring the rate of highway accidents, people could causally link speed limit (cue) to lower accident rates (criterion): “Low speed increases the possibility to respond in time when necessary, and thus the possibility to avoid a collision becomes higher.” Counterintuitively, however, lower speed limits in the present data set are associated with more accidents. That is, causal theories can be and sometimes are wrong—but they nevertheless give rise to intuitions about cue directions.

## Conclusion

The effort–accuracy tradeoff carries the ring of a general law of cognition: Investing less effort is tantamount to achieving lower accuracy. Challenging the belief that accuracy and effort inescapably trade off, research on fast and frugal heuristics has demonstrated that less information and computation can yield better performance. Countering the argument that heuristics’ success rides on the effort put into calculating cue validities and orders, we showed that information limitations that reduce effort do not always hurt accuracy. Simple heuristics can be robust even if simplicity is secured through ordinary information about cue directions, garnered from limited knowledge or found in people’s intuitions.

## References

- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: A process model of risky choice. *Psychological Review*, *113*, 409–432.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74*, 271–280.
- 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*, 63–86.
- Condorcet, N. C. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* [Essay on the application of analysis to the probability of majority decisions]. Paris, France: Imprimerie Royale.
- 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). New York, NY: Oxford University Press.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.
- Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review*, *115*, 119–213.
- García-Retamero, R., Hoffrage, U., & Dieckmann, A. (2007). When one cue is not enough: Combining fast and frugal heuristics with compound cue processing. *Quarterly Journal of Experimental Psychology*, *60*, 1197–1215.
- García-Retamero, R., Takezawa, M., & Gigerenzer, G. (in press). How to learn good cue orders: When social learning benefits simple heuristics. In R. Hertwig, U. Hoffrage, & The ABC Research Group (Eds.), *Social heuristics that make us smart*. New York, NY: Oxford University Press.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.
- Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Heuristics are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas (2008). *Psychological Review*, *115*, 230–239.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
- Gigerenzer, G., Todd, P. M., & The ABC Research Group (Eds.). (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.
- Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*, 494–508.
- Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1191–1206.
- Hogarth, R. M., & Karelaia, N. (2005). Ignoring information in binary choice with continuous variables: When is less “more”? *Journal of Mathematical Psychology*, *49*, 115–124.
- Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*, 733–758.
- Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Market Research*, *40*, 366–371.
- Justin, P., & Persson, M. (2002). PROBABILITIES from EXemplars: A “lazy” algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*, 563–607.
- Katsikopoulos, K. V., & Martignon, L. (2006). Naïve heuristics for paired comparison: Some results on their relative accuracy. *Journal of Mathematical Psychology*, *50*, 488–494.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the “take the best” and “rational” models. *Psychonomic Bulletin & Review*, *11*, 343–352.
- Newell, B. (2005). Re-visions of rationality? *Trends in Cognitive Sciences*, *9*, 11–15.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, England: Cambridge University Press.
- Rakow, T., Hinvest, N., Jackson, E., & Palmer, M. (2004). Simple heuristics from the adaptive toolbox: Can we perform the requisite learning? *Thinking and Reasoning*, *10*, 1–29.
- Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, *112*, 610–628.
- Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, *63*, 129–138.
- Tversky, A. (1972). Elimination of aspects: A theory of choice. *Psychological Review*, *79*, 281–299.
- Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science*, *15*, 307–311.

Received July 1, 2009

Revision received March 18, 2010

Accepted March 19, 2010 ■