PROOF

3
Towards a Rational Theory of Heuristics

Gerd Gigerenzer

Herbert Simon left us with an unfinished task, a theory of bounded rationality. Such a theory should make two contributions. First, it should describe how individuals and institutions *actually make* decisions. Understanding this process would advance beyond as-if theories of maximizing expected utility. Second, the theory should be able to deal with situations of uncertainty where 'the conditions for rationality postulated by the model of neoclassical economics are not met' (Simon, 1989, p. 377). That is, it should extend to situations where one cannot calculate the optimal action but instead has to satisfice, that is, find either a better option than existing ones or one that meets a set aspiration level. This extension would make decision theory particularly relevant to the uncertain worlds of business, investment, and personal affairs.

Simon proposed satisficing as a general alternative to optimizing and also used the term to refer to a specific decision-making heuristic. Consider his account of why he studied political science and economics: 'I simply picked the first profession that sounded fascinating' (Simon, 1978, p. 1). This process is the essence of the *satisficing heuristic*, to set an aspiration level that defines what a satisfactory option would be and then choose the first alternative that meets that level. Satisficing can deal with uncertainty, that is, with situations where not all alternatives and consequences can be foreseen. The same rule is used for business decisions. Developers of high-rise office buildings and malls report that they decide in favor of an investment if they can get at least *x* return in *y* years (Berg, 2014a), and BMW dealers price used cars by setting an aspiration level and lowering it when the car is not sold after about 30 days (Artinger and Gigerenzer, 2015). Satisficing is a heuristic in the *adaptive toolbox* of individuals and organizations, but not the only one. When

34

making his own retirement investments, economist Harry Markowitz did not use his Nobel Prize winning 'optimal' mean-variance model, but a simple heuristic. 'I thought, "You know, if the stock market goes way up and I'm not in it, I'll feel stupid. And if it goes way down and I'm in it, I'll feel stupid,"' Markowitz recalls, 'so I went 50-50' (interview with Bower, 2011, p. 26). An equal split between bonds and equities is an instance of the 1/N heuristic. When wealth is allocated across a menu of N stocks (instead of stocks and bonds, as Markowitz did), studies indicate that 1/N typically outperforms the mean-variance model in the real world of finance, where the assumptions of the mean-variance model are not met (DeMiguel *et al.*, 2009).

Simon himself never systematically studied the heuristics in the adaptive toolbox, nor did he analyze the conditions under which these heuristics are successful – their 'ecological rationality' (Gigerenzer and Selten, 2001). Simon was well aware that he had provided a direction, but not a theory.[1] As he wrote to me shortly before his death, 'I did not want to give the impression that I thought I had "solved" the problem of creating an empirically grounded theory of economic phenomena. What I was trying to do is to call attention for the need for such a theory' (see Gigerenzer, 2004, p. 406). Earlier, he had wondered why his call for realism was received with 'something less than unbounded enthusiasm' and 'largely ignored as irrelevant for economics' (Simon, 1997, p. 269). I believe the answer is that he challenged two profound methodological commitments of neoclassical economists, the twin allegiance to optimization and as-if theories (Berg, 2014b). Going beyond these, Simon called for a shift towards:

1. *Uncertainty:* Analyze decision-making under uncertainty, where the optimal action cannot be determined.
2. *Process:* Design formal theories of the process of decision-making rather than as-if theories.

Let me explain. In this chapter, I use the term *risk* to describe situations in which all alternatives, outcomes, and probabilities are known for sure. The prototype is the choice between monetary gambles where all payoffs are well-defined. Correspondingly, I use the term *uncertainty* for situations where not all is known or can be foreseen. Similar distinctions have been made before. Knight (1921) distinguished between measurable probabilities, that is, frequencies and propensities, and those that cannot be measured empirically: 'a *measurable* uncertainty, or "risk" proper...is so far different from an *unmeasurable*

PROOF

one that it is not in effect an uncertainty at all' (p. 20). L. J. Savage drew a comparable line between small worlds where Bayesian decision theory applies and situations where it does not. For instance, Savage (1972, p. 16) believed it would be 'utterly ridiculous' to apply Bayesian decision theory to problems such as planning a picnic or playing chess, and for different reasons. Planning a picnic, like choosing a profession, is an ill-defined situation, where it is impossible to foresee every possible outcome and where surprises may happen. Thus, the best course of action cannot be calculated in advance. Chess, in contrast, is a well-defined game with an optimal sequence of moves, which, however, no machine or mind can find. In technical terms, the game is computationally intractable – as are most problems that computer scientists work on (Tsotsos, 1991). This differs from tic-tac-toe, where players can easily determine the best sequence of moves, which makes it monotonous for all but small children.

The prototype of an as-if theory is the Ptolemaic model, with the Earth in the center and the planets and the sun orbiting around it in circles and epicycles. Few astronomers believed that planets actually move in such odd-looking epicycles; rather, the theory was designed as a guide for making predictions about planetary positions. Based on Copernicus's heliocentric revolution, Kepler's laws of planetary motion model the actual process of movement, with the planets moving around the sun in ellipses. After Ptolemy's as-if theory was overthrown in favor of a theory of the process, theoretical realism eventually led to better predictions. In the natural sciences, moving from as-if to process is considered progress. Not so in neoclassical economics. In his defense of as-if models, Friedman (1953) argued that the goal is not realism, but prediction. Yet, as I will argue, and as seen in the above example, increased realism is likely to improve prediction.

These two methodological commitments, optimization and as-if, are closely related. The ideal of optimization requires full knowledge of the relevant conditions and thus promotes as-if theories of economic agents who inhabit a world of known risks, not uncertainty. Yet in the real world of business, these risks (such as the space of all possible outcomes and the probability distributions over outcomes and payoffs) are often not known. The standard procedure of neoclassical economics is to transform situations of uncertainty into those of risk in order to be able to determine the best course of action. Whether this optimal solution is actually optimal in the real situation (i.e., under uncertainty) remains up in the air. Consider chess

again, where it is impossible to calculate the optimal sequence of moves although it does exist. In order to optimize, an as-if modeler could reduce the 8x8 board to a 4x4 board with a smaller set of figures. Yet this method would contribute little to mastering the real game. The alternative program is to accept that optimization has its limits and instead focus on analyzing the heuristics that chess masters and computer programs actually use. For some economists, however, a model without optimization does not belong to economics and is therefore inadmissible. Methodological commitments can unite a discipline but may also prove to be a mental straightjacket that inhibits progress.

The Neoclassical Counterrevolution

If we think of Simon's vision as a revolutionary program, the reason why it has been largely ignored can be called a 'neoclassical counterrevolution' supported, surprisingly, by the majority of behavioral economists.

First, neoclassical economists have declared bounded rationality to be nothing but full rationality in disguise. It is nothing new, so the argument goes, and we can therefore ignore it. For instance, in his essay in memory of Herbert Simon, Arrow (2004) insisted that 'boundedly rational procedures are in fact fully optimal procedures when one takes account of the cost of computation in addition to the benefits and costs inherent in the problem as originally posed' (p. 48). In many economists' view, bounded rationality is simply as-if optimization under constraints; Simon's bounds can be modeled by merely adding new constraints, such as those of memory and problem-solving ability, to the standard budget constraints. Simon once told me that he had considered suing colleagues who misused his term *bounded rationality* to refer to optimization.

Second, consider behavioral economics. Simon was one of its creators, but soon dropped out when Kahneman, Tversky, Thaler, and their followers took over and changed its course. Contrary to Simon, these researchers argued that there is nothing wrong with the theory of expected utility maximization but that the fault lies with people who do not follow it. 'Our research attempted to obtain a map of bounded rationality, by exploring the systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models' (Kahneman, 2003, p. 1449).[2] Yet, suboptimal beliefs were not what Simon had in mind; as he points

out, 'bounded rationality is not irrationality' (Simon, 1985, p. 297). Nevertheless, many psychologists have come to believe that bounded rationality is the study of deviations from rationality.

Although behavioral economists started out with the promise of greater psychological realism, most have surrendered to the as-if methodology. Cumulative prospect theory, inequity-aversion theory, and hyperbolic discounting are all as-if theories. They retain the expected utility framework and merely add free parameters with psychological labels (Berg and Gigerenzer, 2010), which is like adding more Ptolemaic epicycles in astronomy. The resulting theories are more unrealistic than the expected utility theories they are intended to improve on. Behavioral economics has largely become a repair program for expected utility maximization.

In sum, both neoclassical and behavioral economists altered and fitted Simon's program of bounded rationality to their programs, emphasizing rationality and irrationality, respectively. Despite this apparent contradiction, both groups regard the classical utility maximization framework as the sole way to model rational decisions. The behavioral economics 'revolution,' as it was once called, has boiled down to defending the neoclassical commitments to optimization and as-if theories.

Adding Parameters to the Utility Function Helps Predict the Past, Not Necessarily the Future

But what is wrong, one might ask, with these commitments, given that they provide a general framework of rationality? The price paid for the lack of realism is predictive power, which is exactly what Milton Friedman held to be the benchmark of a successful theory. Adding more parameters to the utility function leads to a better fit, but may mean losing predictive power. For instance, a study showed that cumulative prospect theory excelled in predicting the past, that is, when its parameters were fitted to known data, but when predicting the future, it did systematically worse than a simple rule called the *priority heuristic* (for hard choices, that is, gambles with similar expected values) and than expected value theory (for easy choices; see Brandstätter *et al.*, 2006). This result is not accidental. Neither the priority heuristic nor expected value theory has free parameters, and thus both avoid error due to parameter estimation, a source of prediction error I consider below. This simplicity can be a strength; the priority heuristic implies the four-fold pattern of risk attitude and other violations of expected utility theory

without needing a new set of parameters for each class of violations (Katsikopoulos and Gigerenzer, 2008). Note that predictive performance is not the same as R^2 in data fitting, which amounts to hindsight; prediction is about foresight, as in out-of-sample prediction when an inference is made from a sample to another sample or a population.

In their review of half a century of research, D. Friedman, Isaac, James, and Sunder (2014) analyzed the empirical evidence for how well Bernoulli functions – such as utility of income functions, utility of wealth functions, and the value function in prospect theory – perform in terms of predictive accuracy. They concluded: 'Their power to predict out-of-sample is in the poor-to-nonexistent range, and we have seen no convincing victories over naïve alternatives' (p. 3). Similarly, Stewart, Reimers, and Harris (2014) experimentally showed that no stable mapping exists between attribute values and subjective equivalents, as assumed in expected utility theories and their modifications, such as prospect theory and hyperbolic discounting theory. This instability was documented long ago in psychophysical research (Brunswik, 1934; Parducci, 1965). If D. Friedman *et al.* (2014) are correct, then expected utility theories and their modifications fail both at describing the process of decision-making and at accurately predicting the actual outcomes.

The Argument: Better Realism, Better Prediction

In the following, I start with Milton Friedman's statement that the measure of a good theory is its predictive power and derive Simon's realism – rather than Friedman's as-if – as a logical consequence. Friedman (1953, p. 41) wrote that a theory should be evaluated 'only by seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories.' I will argue: (1) that simple heuristics can predict better than highly parameterized models under a wide range of conditions; and also (2) model the process of how individuals and organizations actually make decisions; and therefore, (3) that Friedman's goal of prediction implies studying simple heuristics, not only as-if theories. In this way, I derive Simon's call for realism from Friedman's call for good theories.

Specifically, I show that the error in prediction (unlike in data fitting) has two components that we can influence: bias and variance. Prediction by simple heuristics tends to decrease the variance component, while adding free parameters increases it (Gigerenzer and Brighton,

PROOF

2009). Next, I distinguish three ways of reducing error from variance and show that these correspond to three classes of heuristics that humans rely on. Finally, I show that in natural environments, the bias component of error generated by heuristics appears to be surprisingly low. Together, the analysis of bias and variance specifies the conditions for when simple heuristics predict better than complex 'rational' models and provides an explanation of why simple heuristics can be rational.

The Ecological Rationality of Heuristics

In my own work, I have tried to lay the foundations for a theory of bounded rationality. Such a theory addresses not only Simon's descriptive question (how *do* people make decisions?), but also a normative one (how *should* people make decisions *under uncertainty*?). The study of the adaptive toolbox asks the descriptive question: What is the repertoire of heuristics available to an individual or organization? Its methods are experimentation and observation, and the results are algorithmic models of heuristics, such as satisficing and $1/N$. The study of the 'ecological rationality' of heuristics asks the normative question: What are the conditions under which a heuristic leads to a better outcome than a competing strategy? Its methods are analysis and computer simulation, and the results are the conditions under which a class of heuristics is successful according to a metric such as predictive accuracy.

The study of ecological rationality reaches beyond Simon's call for descriptive process models. However, it was inspired by an analogy of Simon's: 'Human rational behavior (and the rational behavior of all physical symbol systems) is shaped by a scissors whose two blades are the structure of the task environment and the computational capabilities of the actor' (Simon, 1990, p. 7). We have fleshed out his analogy into a systematic theory of ecological rationality (Gigerenzer *et al.*, 2011; Gigerenzer and Selten, 2001). The results explain when and why one should rely on simple heuristics in order to make predictions superior to those of highly parameterized models.

Such a theory of bounded rationality is not about human failure. Rather, it explains how and when people can make good decisions by using less information. In what follows, I will deal exclusively with the ecological rationality of heuristics, focusing on the conditions of their predictive power. For general reviews, see Gigerenzer, Todd, and the ABC Research Group (1999); Hertwig, Hoffrage, and the ABC Research Group (2013); Todd, Gigerenzer, and the ABC Research Group (2012).

The Bias–Variance Dilemma

The cause of error is sometimes conceived as

$$\text{Error} = \text{bias} + \varepsilon, \tag{1}$$

where ε is unsystematic noise (mean zero and uncorrelated with bias), and bias is the systematic difference between the (average) prediction a model makes and the true state. For instance, if the true temporal trajectory of a variable is a polynomial of third degree and a linear regression is used to predict the variable, the model has a systematic bias. Equation (1) is implicit in the argumentation of the heuristics-and-biases program (Kahneman, 2011), where a cognitive error is defined in terms of a systematic bias that arises from ignoring information such as base rates. In this view, if the bias is eliminated, rational judgments are obtained. Yet equation (1) is appropriate only in a world of known risks or data fitting, not for making predictions.

Enter prediction. Consider the problem of estimating the true value μ in a population on the basis of random samples. Each of S samples ($s = 1, \ldots, S$) generates an estimate x_s. The variability of these estimates x_s around their mean $\bar{x}$, which is called *variance* in machine learning, is another source of prediction error (Brighton and Gigerenzer, 2012; Geman *et al.*, 1992). The variance component reflects the sensitivity of the predictions to different samples drawn from the same population. Thus the prediction error (the sum of squared error) can be captured in the equation:

$$\textit{Prediction error} = \text{bias}^2 + \text{variance} + \varepsilon, \tag{2}$$

where

> bias $= \bar{x} - \mu,$, that is, the average deviation of the mean of the sample estimates from the true value, and
> variance $= \frac{1}{s} \sum (x_s - \bar{x})^2$, that is, the mean squared deviation of the sample estimates from their mean $\bar{x}$.

Figure 3.1 provides a visual depiction of bias and variance. The bull's eye represents the true value, and each dart the estimate from a sample. Mr. Bias, whose darts landed on the left dartboard, has a systematic bias but low variance. Mr. Variance, whose darts landed on the right dartboard has no bias because the darts line up exactly around the bull's eye. However, his dart throws show considerable variance. Thus, in prediction, two sources of error (ignoring noise) need to be considered, not one.
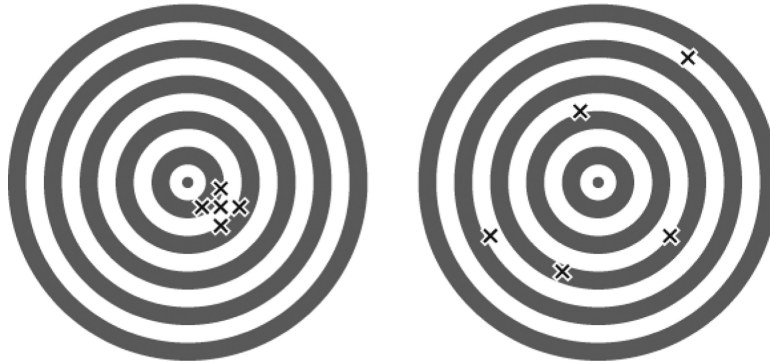
Figure 3.1 A visual depiction of bias and variance
A visual analogy of the two components of prediction error: bias and variance. The bull's eye is the unknown true value μ (here: 0,0) to be predicted. Each dart represents a predicted value x_s based on a random sample from the population with the true value μ. Bias is zero if the mean prediction 'hits' the target. Left: Mr. Bias's strategy results in a systematic bias, whose size is the distance between the mean of the darts thrown and the bull's eye ($\bar{x} - \mu$,), and a low variance, that is, the darts are close together. Right: Mr. Variance's strategy results in zero bias ($\bar{x} = \mu$,), that is, the darts are lined up exactly around the bull's eye, but with considerable variance.

A moderate bias with low variance (left) may lead to better results than would a zero bias with high variance.

The dart analogy in Figure 3.1 does not capture the trade-off between bias and variance. Adding free parameters to a model, which happens when replacing expected utility theory with prospect theory, is likely to reduce the bias component of error, but at the cost of increased variance. By taking away free parameters, such as when replacing expected utility with expected value theory, the opposite happens, a likely reduction in variance at the cost of higher bias. Variance is also influenced by how the parameters are combined, which is determined by the functional form of the model (e.g., multiplicative or exponential). A strategy without any free parameter likely has some bias but no variance. The reason is that the strategy is not sensitive to specific samples and always produces the same prediction. Consider again the 1/N heuristic and the mean-variance model for allocating one's wealth across N stocks. Mean-variance needs to estimate its numerous parameters from stock data and will generate error from both variance and bias. In contrast, 1/N does not estimate any parameters but in fact ignores past data and thus does not generate error from variance but likely from higher bias.
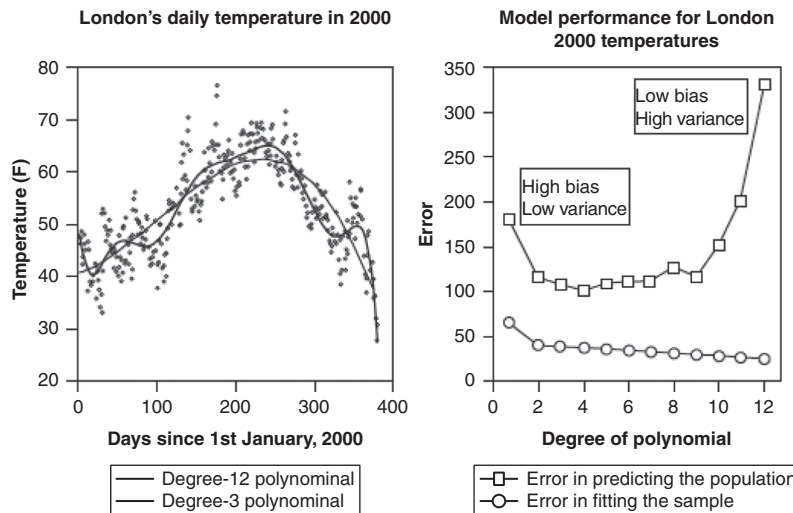
Figure 3.2   Empirical illustration of the bias–variance trade-off when predicting the average daily temperature in London

The bias–variance dilemma in prediction. Left: Data fitting. Each point is the average temperature in London for one of 365 days in the year 2000. The figure shows the best-fitting degree-3 polynomial (thin line) and degree-12 polynomial (thick line), using the least squares method. Clearly, the 12-degree model fits the data best. The error is the sum of squared error. Right: Prediction. The task is to predict the average temperature in London on every day, based on random samples of 30 days. Although the fit increases with higher-degree polynomials (lower curve), the prediction error does not follow this pattern. Rather, there is a u-shaped function between prediction error and complexity of polynomial. For instance, a degree-1 polynomial (i.e., a straight line) generates less error than the degree-12 polynomial, which has less bias but more variance. Adapted from Gigerenzer and Brighton (2009).

Figure 3.2 provides an empirical illustration of the bias–variance trade-off when predicting the average daily temperature in London. The left panel shows the temperature for each day, and a 3-degree and a 12-degree polynomial fitted to the data. The right panel shows that the fit improves (i.e., the error decreases) when the polynomial grows in complexity. A polynomial of degree 364 would guarantee a perfect fit, so that a line can be drawn through each point. But that is not true for prediction. The u-shaped curve in the right panel reveals the trade-off between bias and variance in prediction. Bias is highest for the 1-degree polynomial and lowest for the 12-degree polynomial, while variance shows the opposite pattern. The 4-degree polynomial has the best trade-off between bias and variance, that is, the lowest total error. Note that the 12-degree polynomial, which has the best fit and thus the least bias,

predicts less accurately than the 1-degree polynomial that has a strong bias but less variance.

Let me summarize. The bias–variance dilemma decomposes the total prediction error into bias, variance, and noise. The variance component can be reduced by decreasing the number of parameters and by increasing the sample size. To arrive at good predictions, simpler models predict more accurately to a point, which represents the bias–variance trade-off, while further simplification may lead to an increase in error. Thus, to minimize total error, a certain amount of bias is needed to counteract variance, which is the error due to oversensitivity to characteristics of specific samples. Bias *per se* is not the problem, as assumed in the heuristics-and-biases program. Rather, it can be part of the solution.

Simple Heuristics Can Make Better Predictions

How to Reduce Prediction Error

Consider predicting which of two alternatives will have a higher value on a variable of interest. Assume that the true state of nature can be represented by a linear equation with n attributes (predictors). We do not know what the weights are and want to reduce prediction error due to variance. There are several ways to proceed, each corresponding to a class of simple heuristics that people rely on (Gigerenzer and Gaissmaier, 2011):

1. *One-reason heuristics.* The prediction is based solely on a single predictor among n>1 observable predictors; the other $n - 1$ are ignored. This class of heuristics can be seen as a special case of sequential search heuristics with only one predictor (next class).
2. *Sequential search heuristics.* The prediction is based on a lexicographic rule, that is, if the first predictor does not allow for a decision, then the second is used, and so on. The predictors are ordered by simple correlations between each and the variable of interest, ignoring dependencies (i.e., the covariance structure) among predictors.
3. *Tallying heuristics.* The prediction is based on all n predictors by assigning equal weights to each one and then summing their values.

In each of these classes of heuristics, error due to variance is reduced (relative to a full linear model) because the number of parameters to be estimated is reduced. For instance, the common rationale for all three classes is to avoid the prediction error that results from estimating the $n(n+1)/2$ covariances. Tallying does not estimate the order of the

predictors either, but only their signs (positive or negative). The price for reducing variance is that bias is likely increased (but not necessarily; see below). Consider a few cases.

One-Good-Reason Heuristics

Companies such as catalog retailers, airlines, and hotel chains target their previous customers with product information and special offerings. Not all customers are active, that is, will buy in the future, and predicting which are active is important when reducing marketing costs. The state-of-the-art approach is the Pareto/NBD model and its variants (Schmittlein and Petersen, 1994), where NBD stands for negative binomial distribution. For each previous customer, it yields the probability that he or she is still active, based on the following assumptions (Wübben and von Wangenheim, 2008):

> *Pareto/NBD model:* While the customer is active, purchases follow a Poisson process with purchase rate λ. Customer lifetime is exponentially distributed with dropout rate μ. The purchase rates λ and the dropout rates μ are distributed according to a gamma distribution across the population of customers. The rates λ and μ are distributed independently of each other.

Although this model estimates the probability that a customer is active, it has found little acceptance among experienced managers. Instead, business executives rely on a toolbox of simple heuristics (Verhoef *et al.*, 2002). Wübben and von Wangenheim (2008) observed that managers in an airline and apparel retailer relied on a recency-of-last-purchase (hiatus) rule:

> *Hiatus heuristic:* If a customer has not made a purchase for nine months or longer, classify him/her as inactive, otherwise as active.

The hiatus heuristic is an instance of the class of one-reason heuristics. It considers the hiatus only and ignores other information used by the Pareto/NBD model, such as the number of purchases made. Given that the hiatus heuristic uses only a subset of the relevant information used by the Pareto/NBD model and does not estimate any parameters (if the hiatus is fixed), it might appear to be second best. Equation (2) shows the mistake behind that assumption. The real question is whether the total error that the Pareto/NBD model generates is higher or lower than that of the hiatus heuristic. Wübben and von Wangenheim (2008) put
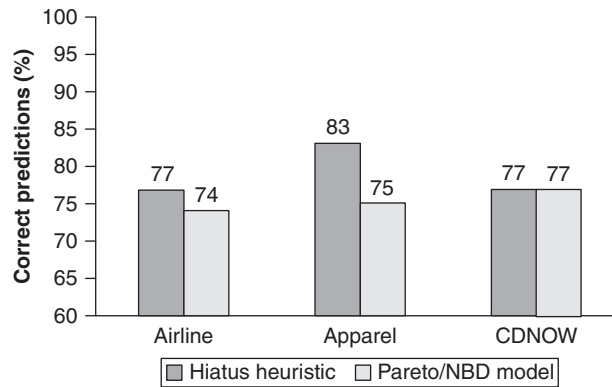
Figure 3.3   Hiatus heuristic made more correct predictions than the Pareto/NBD model

How to predict which customers will buy in the future? Shown is a competitive test of the Pareto/NBD model from marketing science against the hiatus heuristic managers rely on. The heuristic better predicts customer behavior for the airline and the apparel business, and equally well for the CDNOW retailer. With a fixed hiatus (such as 9 months), the heuristic has no free parameter and thus does not make errors due to variance. Note that the heuristic uses only a subset of the data the Pareto/NBD model uses, that is, makes better predictions with less effort. Adapted from Wübben and von Wangenheim (2008).

the issue to an empirical test. They calibrated the Pareto/NBD model to the customer databases of three companies, using 40 weeks of data, and used this calibrated model to predict the next 40 weeks of activities. The third company was the online CD retailer CDNOW, using a six-month hiatus. Figure 3.3 shows that the hiatus heuristic made more correct predictions than did the Pareto/NBD model for the airline customers, with 77% versus 74%, and for the apparel customers, with 83% versus 75%, and matched the number of correct predictions for the CDNOW customers. Less information can be more.

With a fixed hiatus, the hiatus heuristic has bias, but no variance. The Pareto/NBD model likely has less bias, but additional error due to variance because it needs to estimate four parameters from the sample data. The success of the heuristic suggests that its error due to bias is less than the total error made by the Pareto/NDP model.

Sequential Search Heuristics

The *take-the-best heuristic* was the first heuristic that the ABC Research Group systematically studied (Gigerenzer and Goldstein, 1996). It helps decision makers choose between two alternatives based on *n* attributes,

not only one, as with the hiatus heuristic. It orders attributes ($i = 1, \ldots, n$) by their simple validities v_i defined as the proportion of correct decisions c_i:

$$v_i = c_i / t_i \tag{3}$$

The denominator t_i gives the number of cases where the values of the two alternatives on the attribute i differ. If the values on the first attribute do not differ, the next is looked up, and so on in a lexicographic way. The decision is based on the first attribute that differentiates; all other attributes are ignored. Note that their order is determined by simple validities v_i – unlike beta weights in multiple regression, which require estimating the covariance matrix – and that these validities are estimated from samples. For simplicity, I assume here that there is a positive correlation between each attribute and the outcome (dependent) variable. Take-the-best entails three steps:

Search rule: Look through predictors in the order of their validity.
Stopping rule: Stop search when the first predictor is found where the values of the two alternatives differ.
Decision rule: Predict that the alternative with the higher predictor value has the higher value on the outcome variable.

How well does take-the-best predict compared to multiple regression? Figure 3.4 shows the results for 20 prediction tests on economic, demographic, and societal questions, such as which of two houses will have a higher selling price, or which school will have a higher drop-out rate (Czerlinski *et al.*, 1999). In every test, half of the data points were used to fit the parameters, and the other half was predicted, a procedure known as cross validation. On average, multiple regression had the better fit, but take-the-best predicted better. To excel in fitting and fail in prediction is known as overfitting.

The take-the-best heuristic was also more frugal than regression, requiring, on average, only 2.4 predictors compared to 7.7 for regression. Like the hiatus heuristic and the Pareto/NBD model, take-the-best used only a subset of the information used by multiple regression, which protected against estimation error from variance.

Tallying Heuristics

Unlike take-the-best, tallying relies on all predictors but uses equal weights. Figure 3.4 shows that, on average, tallying predicted better
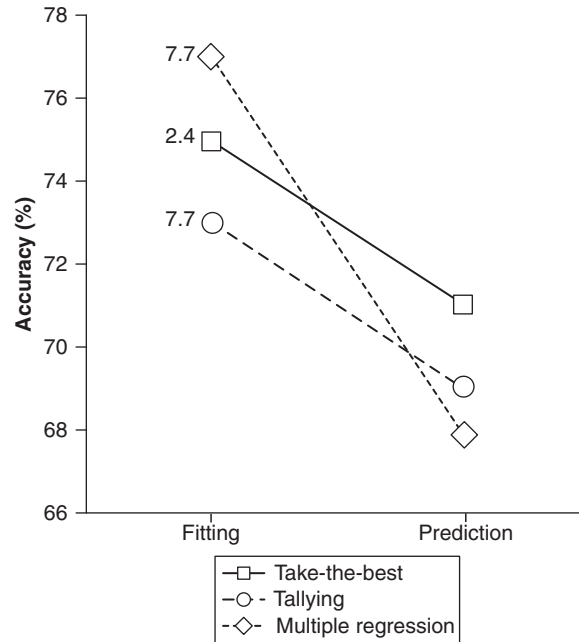
PROOF

Figure 3.4 Results for 20 prediction tests on economic, demographic, and societal questions

Less-is-more effects across 20 prediction tasks in economics, business, biology, and other fields. Two heuristics, take-the-best and tallying, are tested competitively against multiple regression (Czerlinski *et al.*, 1999). Note that many of the tasks are taken from textbooks on regression. The number of attributes ranged between 3 and 18, and the number of alternatives between 11 and 395. Take-the-best orders attributes (predictors) in a simple way (without analyzing dependencies between cues) and uses only the first cue that differentiates between the alternatives. Tallying introduces a different bias; it uses all attributes that regression uses but does not estimate their weights, instead using the same weight for each. Prediction is tested by letting the three strategies estimate their parameters from half of the data set and then testing performance on the other half (cross-validation). Multiple regression estimates beta weights, take-the-best estimates only the order of cues, and tallying only the sign of the cues. By introducing bias, both heuristics make more accurate predictions than regression. For comparison, the performance in fitting data is shown. Regression excels in data fitting because it has more free parameters, but makes fewer correct predictions. Adapted from Gigerenzer and Brighton (2009).

than multiple regression, although not as well as take-the-best. Similarly, Åstebro and Elhedli (2006) used a version of tallying to forecast the commercial success for early-stage ventures and reported that the heuristic predicted 86% of successes and failures correctly, compared to a log-linear regression model that predicted 79% correctly.

Two Misconceptions

These results demonstrate the bias–variance trade-off in the real world of prediction. They also help to correct two widespread misconceptions about heuristics. First, a common explanation for why people rely on heuristics is the accuracy–effort trade-off: Heuristics reduce effort but pay for this with less accuracy (Conlisk, 1996). The effort is often called deliberation costs. Although such a general trade-off sounds plausible, it is incorrect. Heuristics indeed reduce effort, but that does not necessarily reduce accuracy, as the empirical results in this section demonstrate. More generally, the bias–variance dilemma implies that there is no general accuracy–effort trade-off. It also explains when and why higher accuracy results from less effort. These situations are known as less-is-more effects.

A second misunderstanding is the claim that the study of heuristics is unnecessary because a heuristic can always be rewritten as a special case of the general linear model. Indeed, take-the-best and tallying can (Martignon and Hoffrage, 2002), but rewriting does not help to make better predictions. By generalizing a heuristic to a linear model, one can actually lose predictive power by creating more error due to variance, as Figure 3.4 illustrates. After all, rewriting the law of falling bodies as a general polynomial does not add to understanding physics.

Empirical Evidence

There is a large body of empirical studies showing that the classes of heuristics described in the previous sections are good models for how people make decisions, and that people tend to use them in an adaptive way, that is, in situations where they are ecologically rational. For instance, sequential search heuristics, such as in take-the-best, have been studied extensively in both laboratory experiments (e.g., Bröder, 2012; Bergert and Nosofsky, 2007) and 'in the wild' (Gigerenzer *et al.*, 2012). This research showed that decisions made by experts – from airport customs officers to police officers to professional burglars – are best predicted by take-the-best or similar lexicographic rules, while novices and undergraduates try to consider all *n* attributes (Pachur and Marinello, 2013). Hence, experts appear to know intuitively how to reduce error due to variance, which makes their decisions more efficient.

Simon started with the concept of satisficing. Today, we have a large body of empirical evidence for other classes of heuristics, together with formal models of them (Gigerenzer and Gaissmaier, 2011). These formal

models are a scientific leap from the premathematical period of vague labels such as 'availability' and 'System 1' (Kahneman, 2011). Formal models can make testable predictions and enable the normative study of their ecological rationality.

Environmental Structures and Bias

So far, we have focused on variance. But, according to the bias–variance dilemma, the crucial question is how much bias a heuristic produces by reducing variance. To assess bias, one needs to compare predicted outcomes to actual outcomes. Assume again that the actual outcome can be represented by a linear equation with n predictors. Consider again a choice between objects A and B, based on n predictors, where the value of the ith predictor is represented by x_i and weighted in the linear payoff function by w_i. To simplify, assume that the predictors are binary and the weights are nonnegative. The class of strategies we consider are sequential search (lexicographic) heuristics, with one-reason decision-making as a special case with only one predictor.

Environmental Structures

The term *environment* refers to the alternatives, outcomes, payoffs, and all other factors in the model exogenous to the decision maker. Simple environments may be described as a joint distribution of predictors and outcome variables, which induce payoff distributions that depend on actions in the decision maker's choice set. With this broad interpretation of *environment*, we can then investigate which environmental structures 'help' lexicographic heuristics perform well so that they have a small or even zero bias (in addition to small variance). We know of three structural features: noncompensatoriness, dominance, and cumulative dominance.

Noncompensatoriness. The weights w_1, w_2, $w_3, \ldots w_n$ are noncompensatory if they satisfy the $n-1$ inequality constraints:

$$w_i > \sum_{j=i+1}^{k} w_j, i = 1, 2, \ldots, n-1. \tag{4}$$

An example is the set of weights {1, ½, ¼, ⅛} (see Figure 3.5). If the weights are noncompensatory, then a linear rule (with the same order of predictors) will always lead to the same choice as a lexicographic rule (Martignon and Hoffrage, 2002). Take the example of weights above.
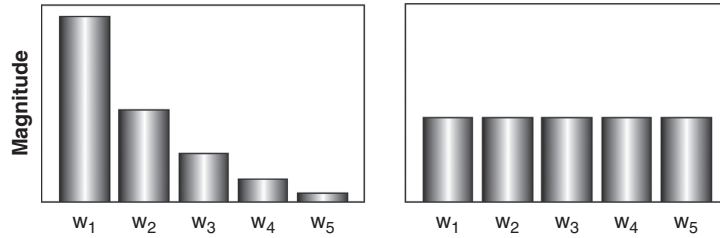
Figure 3.5 With a fixed hiatus, the hiatus heuristic has bias, but no variance
Left: A graphical illustration of noncompensatoriness. If the weights of a linear rule are non-compensatory, such as 1, ½, ¼, ⅛, and ⅟₁₆, then a lexicographic heuristic with the same order of attributes will always make the same prediction as the linear rule. Therefore, if the true state of nature can be represented by a linear rule and noncompensatoriness holds, a lexico-graphic heuristic has no bias. *Right*: For comparison, a set of weights where tallying has no bias. Adapted from Martignon and Hoffrage (2002).

If the lexicographic rule yields decisions on the basis of the first predictor (with weight 1), every linear rule will match this choice, because the sum of all other weights (½+¼+...) will always be smaller than the weight of the first predictor. Thus, if the true state of nature is linear and noncompensatoriness holds, then a lexicographic heuristic with the same order of predictors has no bias.

Dominance. If alternative *A* has a value higher than or equal to alternative *B* in all *n* predictors, and a higher value on at least one predictor, then alternative *A* dominates alternative *B*. Figure 3.6 (top) illustrates dominance. If *A* dominates *B*, a lexicographic heuristic (and tallying) will arrive at the same prediction as a linear rule. In terms of a linear rule, dominance means that all differences $w_i \Delta x_i = w_i(x_{iA} - x_{iB})$ are nonzero and at least one is positive; thus, the linear rule chooses *A* over *B*. This result holds for any (nonnegative) weights and does not depend on noncompensatory ones. Thus, if the true state of nature is linear and dominance holds, a lexicographic heuristic has no bias.

Cumulative dominance. The cumulative profile of an alternative consists of *n* values, where the *i*th value is the sum of the first *i* values. Alternative *A* cumulatively dominates *B* if its cumulative profile exceeds or equals the cumulative profile of *B* in every term and exceeds it in at least one term (Bauscells *et al.*, 2006). If cumulative dominance holds, then a linear rule (with the same order of predictors) predicts the cumulative dominant object, just as a lexicographic rule does. Consider the example in Figure 3.6 (bottom). Unlike in the top panel, dominance does not hold. To check for cumulative dominance, one first compares
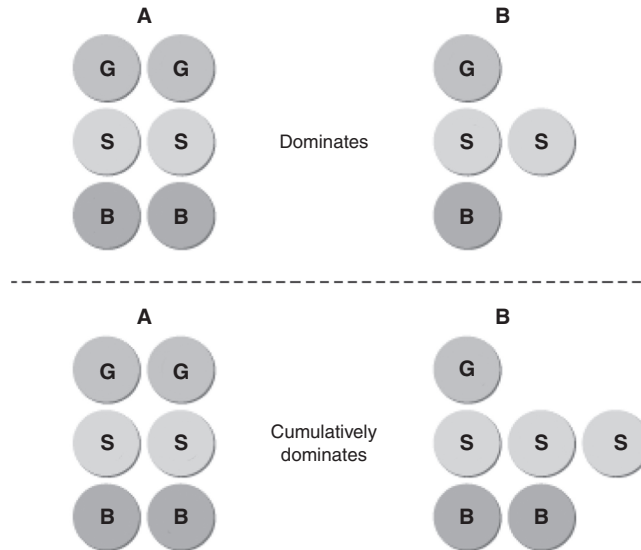
Figure 3.6 Cumulative dominance
A pictorial illustration of dominance and cumulative dominance. Which of two alternatives, *A* and *B*, is more valuable? The alternatives vary on three attributes, gold, silver, and bronze coins. In the top panel, *A* dominates *B* because it has more gold and bronze coins, and as many silver coins. In the bottom panel, dominance does not hold, but cumulative dominance does. To check for cumulative dominance, one first compares the number of gold coins, then the number of gold and silver coins, and finally the number of all coins. In every comparison, alternative *A* has at least as many coins as *B*, and more coins in one. It has more gold coins, an equal number of gold and silver coins, and an equal number of gold, silver, and bronze coins. Thus, *A* cumulatively dominates *B*. If dominance or cumulative dominance holds, a linear model makes the same choice (alternative A) as a lexicographic heuristic with the same order of cues. Adapted from Şimşek (2014).

A and *B* on the top attribute (gold), *A* has two gold coins, and *B* only one. Then *A* and *B* are compared on the sum of the top two attributes; here the number of coins is the same. Finally, the comparison is made on all three attributes, and again there is no difference. Because there is always zero difference and one difference in favor of *A*, *A* cumulatively dominates *B*.

Thus, unlike simple dominance, cumulative dominance requires an order of predictors or attributes. The cumulative difference can be defined as:

$$\Delta x_i' = \sum_{j=1}^{i} \Delta x_j. \tag{5}$$

If the weights $w_1, \ldots, w_k$ are positive and decreasing, then one needs to check only for cumulative dominance. Otherwise, alternative *A* cumulatively dominates *B* if all terms $w_i' \Delta x_i'$ are nonnegative and at least one of them is positive, where $w_i' = w_i - w_{i+1}, i = 1, 2, \ldots, n-1$, and $w_n' = w_n$. Note that dominance implies cumulative dominance, but not vice versa.

These three environmental conditions influence the bias component of the error. Noncompensatoriness refers to the relative strength of the predictors in the environment, while the two dominance conditions refer to the relative quality of alternatives (Katsikopoulos, 2011). In sum, if either noncompensatoriness or dominance or cumulative dominance holds, then lexicographic heuristics will have no bias, relative to a linear rule.

How Often Do These Conditions Hold in the Real World?

It is not easy to answer this question because there is no way to define the set of all prediction problems and draw a random sample. But it is possible to investigate a large and diverse number of natural data sets. Şimşek (2013) analyzed 51 data sets from online repositories, textbooks, research publications, packages for R statistical software, and individual scientists collecting field data. The data sets spanned content areas as diverse as biology, business, computer science, ecology, economics, education, engineering, and medicine, among others. The number of attributes ranged from 3 to 21, which were numeric or binary; the number of objects (alternatives) from 12 to 601, corresponding to numbers of possible pairwise comparisons ranging from 66 to 180,300. Each of these comparisons amounts to a prediction being made.

How often was one or more of the three conditions – noncompensatoriness, dominance, and cumulative dominance – satisfied? The result was surprising. The median for the 51 data sets was 90% (Şimşek, 2013). That is, in half of the data sets, more than 90% of the decisions encountered were such that a lexicographic rule yielded the same prediction as a linear model. When the predictors were dichotomized at the median, this number increased to 97%. In other words, in the majority of the cases, the lexicographic heuristics had the same bias as a linear model. Together with the potential for reducing variance, this result explains why simple heuristics often outperform linear models in prediction, as shown in Figure 3.4.

In summary, the prediction error has two primary components that can be influenced, bias and variance. Variance can be decreased by decreasing the number (and kind) of free parameters and by increasing

the sample size. Simple heuristics decrease variance because they have few or zero free parameters. To analyze the bias component, one needs to know the true process that generates actual outcomes, which was assumed to be a linear function of *n* predictors or attributes whose order is known. Under those assumptions, three environmental conditions were described that guarantee that lexicographic heuristics have no bias when choosing between two alternatives: noncompensatoriness, dominance, and cumulative dominance. Alternatively, if the true process is unknown, the bias is equivalent to that of a linear model. An analysis of a diverse collection of data showed that one or more of these conditions were in place in 90% (97% for binary attributes) of the cases. Together, these results provide an explanation of less-is-more effects, that is, situations where using a subset of the available information (e.g., ignoring all data except the hiatus) leads to better predictions than using all available information.

Methodological Principles

Finally, the study of the adaptive toolbox and that of ecological rationality entail adherence to three methodological principles:

1. *Algorithmic models of heuristics*, not verbal labels (such as availability, System 1, and near-tautologies; see Gigerenzer, 1996; Gigerenzer *et al.*, 2012).
2. *Tests of prediction*, not fitting data.
3. *Competitive testing* (such as testing the predictions of two fully specified models), not testing of a single model.

For instance, the research on the hiatus heuristic uses an algorithmic model of the heuristic, tests its performance in prediction, and compares it to the Pareto/NBD model. Although these methodological principles should be obvious they are not widely followed in those parts of the behavioral economics literature that propose verbal labels, rather than algorithmic models of heuristics, or fit models, such as prospect theory, to data sets without out-of-sample prediction and without testing them competitively against models of simple heuristics.

Conclusion

In this chapter, I started with Milton Friedman's dictum that the measure of a good theory is its predictive power and derived Simon's realism,

rather than as-if, as the logical consequence. Specifically, I showed that the error in prediction (unlike in data fitting) has two components, bias and variance. The use of simple heuristics likely decreases variance, while the use of models with many free parameters tends to increase it. Moreover, an analysis of natural environments indicates that simple heuristics generate a surprisingly small bias, relative to linear models. Investigating the balance between bias and variance allows for deriving the conditions under which simple heuristics can predict more accurately than complex 'rational' models, and vice versa. These results can unite Friedman's dictum with Simon's call for realism. If rationality means making better predictions, we should seriously investigate the adaptive toolbox of humans and the ecological rationality of heuristics rather than adding more parameters to as-if utility functions.

As a consequence, the use of simple heuristics by economic agents should not be routinely attributed to mere deliberation costs or even irrationality. Instead, it should be recognized that some degree of bias actually enables better performance in situations of uncertainty. Risk and uncertainty can require different sets of tools, of statistics, and of heuristics. In Simon's words (1981, p. 36), uncertainty 'places a premium on robust adaptive procedures instead of strategies that work well only when finely tuned to precisely known environment.'

Many years ago, a well-known behavioral economist told me with utmost conviction: 'Look, either reasoning is *rational* or it's *psychological*.' This false opposition between what is regarded as rational versus psychological has haunted me since. It is time to rethink the nature of rationality. A theory of bounded rationality based on the twin foundations of the adaptive toolbox and ecological rationality can be a start. Pursuing this goal is a step towards making progress on the unfinished task Simon left behind and may even contribute to unifying economics and psychology.[3]

Notes

1. Simon developed his thinking over decades, a complex process that I cannot give due justice to in this article. For instance, he introduced a satisficing heuristic in 1955, but in the Appendix of that article he presented an optimization model that maximizes expected value of the sales price, similar to the optimization under constraints model that Stigler (1961) later proposed. In the introduction to this article in his collected papers, Simon (1979, p. 3) made it clear that he thinks of satisficing as nonoptimizing: 'Satisficing, aiming at the good when the best is incalculable, is the key device.' Similarly, in early writings he sometimes linked bounded rationality to cognitive

limitations, while later he linked it to cognition and environment – the scissors analogy – and argued that it is impossible to understand behavior by looking at the one blade of cognitive limitations only. Thus, part of the misinterpretations of Simon's concept of bounded rationality I point out later may be due to his own development in thinking.

2. Although Simon's bounded rationality is commonly presented as a forerunner of Kahneman's heuristics and biases program, the latter's relation to bounded rationality appears to be an afterthought. In fact, Simon is not cited at all in Kahneman and Tversky's major early papers (all reprinted in Kahneman, Slovic, & Tversky, 1982). Simon is briefly mentioned in the preface to this anthology, apparently more as a nod to a distinguished figure than an acknowledgment of a significant intellectual debt (Lopes, 1992).

3. For helpful comments I would like to thank Florian Artinger, Nathan Berg, Henry Brighton, Ralph Hertwig, Perke Jacobs, Konstantinos Katsikopoulos, Shenghua Luan, Thorsten Pachur, and Ozgür Simsek.

References

Artinger, F., and Gigerenzer, G. (2015). Aspiration-Adaptation,P price Setting, and the Used-Car Market. Unpublished manuscript.

Arrow, K. J. (2004). Is Bounded Rationality Unboundedly Rational? Some Ruminations. In *Models of a Man: Essays in Memory of Herbert A. Simon*, ed. M. Augier and J. G. March. Cambridge, MA: MIT Press.

Åstebro, T. and Elhedhli, S. (2006). The Effectiveness of Simple Decision Heuristics: Forecasting Commercial Success for Early-Stage Ventures. *Management Science*, 52, 395–409.

Baucells, M., Carrasco, J. A., and Hogarth, R. M. (2006). *Cumulative Dominance and Heuristic Performance in Binary Multi-Attribute Choice*. Available at SSRN: http://www.researchgate.net/publication/23695656_Cumulative_Dominance_ and_Heuristic_Performance_in_Binary_Multi-Attribute_Choice.

Berg, N. (2014a). Success from Satisficing and Imitation: Entrepreneurs' Location Choice and Implications of Heuristics for Local Economic Development. *Journal of Business Research,* 67, 1700–09. doi: 10.1016/j.jbusres.2014.02.016

Berg, N. (2014b). The Consistency and Ecological Rationality Schools of Normative Economics: Singular versus Plural Metrics for Assessing Bounded Rationality. *Journal of Economic Methodology, 21,* 375–95.

Berg, N. and Gigerenzer, G. (2010). As-if Behavioral Economics: Neoclassical Economics in Disguise? *History of Economic Ideas, 18*, 133–65. doi: 10.1400/140334

Bergert F. B. and Nosofsky, R. M. (2007). A Response-Time Approach to Comparing Generalized Rational and Take-the Best Models of Decision Making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 331, 107–29.

Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The Priority Heuristic: Making Choices without Trade-offs. *Psychological Review*, 113, 409–32. doi: 10.1037/0033-295X.113.2.409

Brighton, H. and Gigerenzer, G. (2012). Are Rational Actor Models 'Rational' Outside Small Worlds? In: *Evolution and Rationality: Decisions, Co-operation and Strategic Behavior*, ed. S. Okasha and K. Binmore. Cambridge: Cambridge University Press, 84–109.

PROOF

Bower, B. (2011). Simple Heresy. Rules of Thumb Challenge Complex Financial Analyses. *ScienceNews*, 179, 26.

Bröder, A. (2012). The Quest for Take-the-Best. In *Ecological Rationality: Intelligence in the World*, ed. P. M. Todd, G. Gigerenzer and the ABC Research Group. New York: Oxford University Press, 216–40.

Brunswik, E. (1934). *Wahrnehmung und Gegenstandswelt: Grundlegung einer Psychologie vom Gegenstand her.* Leipzig: Deuticke.

Czerlinski, J., Gigerenzer, G. and Goldstein, D. G. (1999). How Good are Simple Heuristics? In: *Simple Heuristics that Make Us Smart*, ed. G. Gigerenzer, P. M. Todd, and the ABC Research Group. New York: Oxford University Press, 97–118.

Conlisk, J. (1996). Why Bounded Rationality? *Journal of Economic Literature*, 34, 669–700.

DeMiguel V, Garlappi L, and Uppal R. (2009). Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22, 1915–53.

Friedman, M. (1953). *Essays in Positive Economics.* Chicago: University of Chicago Press.

Friedman, D., Isaac, R. M., James, D. and Sunder, S. (2014). *Risky Curves. On the Empirical Failure of Expected Utility.* New York: Routledge.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4, 1–58.

Gigerenzer, G. (1996). On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky. *Psychological Review*, 103, 592–6.

Gigerenzer, G. (2004). Striking a Blow for Sanity in Theories of Rationality. In: *Models of a Man: Essays in Memory of Herbert A. Simon*, ed. M. Augier and J. G. March. Cambridge, MA: MIT Press, 389–409.

Gigerenzer, G., and Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, 1, 107–43.

Gigerenzer, G., Fiedler, K. and Olsson, H. (2012). Rethinking Cognitive Biases as Environmental Consequences. In *Ecological Rationality: Intelligence in the World*, ed. P. M. Todd, G. Gigerenzer and the ABC Research Group. New York: Oxford University Press, 80–110.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, 62, 451–82.

Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review, 103*, 650–69. doi: 10.1037/0033-295X.103.4.650

Gigerenzer, G., Hertwig, R. and Pachur, T. (eds) (2011). *Heuristics: The Foundations of Adaptive Behavior.* New York: Oxford University Press.

Gigerenzer, G., and Selten, R. (eds). (2001). *Bounded Rationality: The Adaptive Toolbox.* Cambridge, MA: MIT Press.

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

Hertwig, R., Hoffrage, U. and the ABC Research Group. (2013). *Simple Heuristics in a Social World.* New York: Oxford University Press.

Kahneman, D. (2003). Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice. In *Les Prix Nobel: The Nobel Prizes 2002*, ed. T. Frangsmyr. Stockholm: Nobel Foundation, 1449–89.

Kahneman, D. (2011). *Thinking Fast and Slow.* London: Allen Lane.

PROOF

Kahneman, D., Slovic, P. and Tversky, A. (eds). (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.

Katsikopoulos, K. V. (2011). Psychological Heuristics for Making Inferences: Definition, Performance, and the Emerging Theory and Practice. *Decision Analysis*, 8, 10–29.

Katsikopoulos, K. V. and Gigerenzer, G. (2008). One-Reason Decision-making: Modeling Violations of Expected Utility Theory. *Journal of Risk and Uncertainty*, 37, 35–56.

Knight, F. (1921). *Risk, Uncertainty and Profit* (Vol. XXXI). Boston: Houghton Mifflin.

Lopes, L. L. (1992). Three Misleading Assumptions in the Customary Rhetoric of the Bias Literature. *Theory & Psychology*, 2, 231–6. doi: 10.1177/0959354392022010

Martignon, L. and Hoffrage, U. (2002). Fast, Frugal, and Fit: Lexicographic Heuristics for Paired Comparison. *Theory and Decision*, 52, 29–71. doi: 10.1023/A:1015516217425

Pachur, T. and Marinello, G. (2013). Expert Intuitions: How to Model the Decision Strategies of Airport Customs Officers? *Acta Psychologica*, 144, 97–103.

Parducci, A. (1965). Category Judgment: A Range-Frequency Model. *Psychological Review*, 72, 407–18.

Savage, L. J. (1972). *The Foundations of Statistics* (2nd edn). New York: Wiley.

Schmittlein, D. C., and Peterson, R. A. (1994). Customer Base Analysis: An Industrial Purchase Process Application. *Marketing Science*, 13, 41–67. doi: 10.1287/mksc.13.1.41

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics,* 69, 99–118. doi: 10.2307/1884852

Simon, H. A. (1978). Rationality as Process and as Product of Thought. *American Economic Review*, 68, 1–16.

Simon, H. A. (1979). *Models of Thought.* New Haven: Yale University Press.

Simon, H. A. (1981). *The Sciences of the Artificial* (2nd edn). Cambridge, MA: MIT Press.

Simon, H. A. (1985). Human Nature in Politics: The Dialogue of Psychology with Political Science. *American Political Science Review*, 79, 293–304.

Simon, H. A. (1989). The Scientist as Problem Solver. In *Complex Information Processing: The Impact of Herbert A. Simon*, ed. D. Klahr and K. Kotovsky. Hillsdale, NJ: Elbaum, 375–98.

Simon, H. A. (1990). Invariants of Human Behavior. *Annual Review of Psychology*, 41, 1–19. doi: 10.1146/annurev.ps.41.020190.000245

Simon, H. A. (1997). *Models of Bounded Rationality, Volume 3: Empirically Grounded Economic Reason*. Cambridge, MA: MIT Press.

Şimşek, Ö. (2013). Linear Decision Rule as Aspiration for Simple Decision Heuristics. In *Advances in Neural Information Processing Systems: Vol. 26: 27th Annual Conference on Neural Information Processing Systems 2013 [online version]*, ed. C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K.Q. Weinberger. Red Hook, NY: Curran Associates, 2904–12.

Stewart, N., Reimers, S. and Harris, A. J. L. (2014). On the Origin of Utility, Weighting, and Discounting Functions: How They Get theirShapes and How to Change their Shapes. *Management Science, 61*, 687–705. doi: 10.1287/mnsc.2013.1853

Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy*, 69, 213–25. doi: 10.1086/258464

Todd, P. M., Gigerenzer, G. and the ABC Research Group. (2012). *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press.

Tsotsos, J. (1991). Computational Resources do Constrain Behavior. *Behavioral Brain Sciences,* 14, 506–7.

Verhoef, P. C, Spring, P. N., Hoekstra, J. C. and Leeflang, P. S. H. (2002). The Commercial Use of Segmentation and Predictive Modeling Techniques for Database Marketing in the Netherlands. *Decision Support Systems*, 34, 471–81.

Wübben, M. and von Wangenheim, F. (2008). Instant Customer Base Analysis: Managerial Heuristics Often 'Get It Right.' *Journal of Marketing*, 72, 82–93.