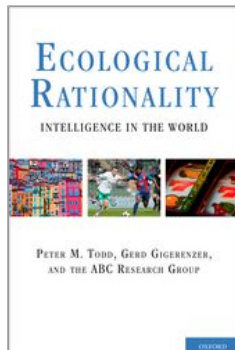


How Estimation Can Benefit From an Imbalanced World

University Press Scholarship Online

Oxford Scholarship Online



Ecological Rationality: Intelligence in the World

Peter M. Todd and Gerd Gigerenzer

Print publication date: 2012

Print ISBN-13: 9780195315448

Published to Oxford Scholarship Online: May 2012

DOI: 10.1093/acprof:oso/9780195315448.001.0001

How Estimation Can Benefit From an Imbalanced World

Ralph Hertwig

Ulrich Hoffrage

Rüdiger Sparr

DOI:10.1093/acprof:oso/9780195315448.003.0116

Abstract and Keywords

This chapter analyzes how valuable the assumption of systematic environment imbalance is for performing rough-and-ready intuitive estimates, which people regularly do when inferring the quantitative value of an object (e.g., its frequency, size, value, or quality). The chapter outlines how systematic environment imbalance can be quantified using the framework of power laws. It investigates to what extent power-law characteristics and other statistical properties of real-world environments can be allies of two simple estimation heuristics, QuickEst and the mapping heuristic. The analyses, which involve comparing the estimation performances of the heuristics relative to more complex strategies, demonstrate that QuickEst could be particularly suited for deriving rough-and-ready estimates in skewed distributions with highly

How Estimation Can Benefit From an Imbalanced World

dispersed cue validities, whereas the mapping heuristic might be most suited when the cues have similar validities.

Keywords: power law, estimation, statistical properties, QuickEst heuristic, mapping heuristic, skewness

Both organism and environment will have to be seen as systems, each with properties of its own, yet both hewn from basically the same block.

Egon Brunswik

Much of the world is in a state of *predictable imbalance*. This is a notion that is commonly attributed to the Italian economist Vilfredo Pareto, who was a professor of political economy at the University of Lausanne in Switzerland in the 1890s. He first introduced what is now known as the *Pareto law* of income distribution in his *Cours d'Économie Politique* (Pareto, 1897) where he described the finding that income and wealth distributions exhibit a common and specific pattern of imbalance across times and countries. In qualitative terms, the predictable imbalance in income and wealth distributions is that a relatively small share of the population holds a relatively large share of the wealth.

For an illustration, let us turn to the exclusive circle of the global rich. Each year, *Forbes* magazine publishes its famous annual ranking of the wealthiest people around the globe. The 2008 listing included a total of 1,125 billionaires, among them not only the “usual suspects” such as Bill Gates and Warren Buffett, but also newcomers such as Mark Zuckerberg, founder of the social networking site *Facebook*, and at age 23 years possibly the youngest self-made billionaire ever (Kroll, 2008). Even in this highly selective group of the world’s super-rich, the distribution of wealth is highly unbalanced. One measure of this imbalance is the share of the collective net worth of these wealthiest people that goes to the top 1% of them. In 2008, the 11 richest billionaires’ collective fortune amounted to as much as that of the 357 “poorest” billionaires.

(p.380) One consequence of this predictable imbalance is that if somebody were to estimate the net worth of a billionaire, say, Donald Trump, a good starting point would be to assume

How Estimation Can Benefit From an Imbalanced World

that the fortune in question is modest. Why? Because most billion-dollar fortunes in this skewed world of incomes and wealth are small.

The goal of this chapter is to analyze how valuable the assumption of systematic environment imbalance is for performing rough estimates. By such estimates, we mean the routine assessment of quantities (e.g., frequencies, sizes, amounts) in which people regularly engage when they infer the quantitative value of an object (such as its frequency, size, value, or quality). To this end, we first outline how systematic environment imbalance can be described using the framework of power laws. Then, we investigate to what extent power-law characteristics as well as other statistical properties of real-world environments can be allies of simple heuristics in performing rough-and-ready estimates, thereby leading to ecological rationality. For this purpose we will introduce two heuristics: The first, QuickEst, uses simple building blocks for ordered cue search and stopping and is particularly suited for skewed environments. The second, the mapping model or mapping heuristic, is built on the simplifying decision mechanism of tallying and can be applied to a broader range of distributions.

The Ubiquity of Power-Law Regularities

The Pareto law belongs to the family of power laws. A power-law distribution of the sizes of objects (on some dimension) implies a specific relationship between the rank of an object and its size. Let us illustrate this relationship with a graph (adopting Levy & Solomon's, 1997, approach to analyze power-law distribution of wealth). Suppose one takes all the billionaires in the *Forbes* 2008 (Kroll, 2008) listing, ranks them by their wealth, and then plots the billionaires' wealth against their rankings. Figure 15-1a shows the resulting J-shaped distribution (where the "J" is rotated clockwise by 90 degrees), which reveals that a great many billionaires have "small" fortunes, and only very few have resources much greater than those small fortunes. This picture becomes even more interesting if it is redrawn with logarithmic horizontal and vertical axes. As Figure 15-1b shows, the resulting *rank-size distribution* (Brakman, Garretsen, Van Marrewijk, & van den

How Estimation Can Benefit From an Imbalanced World

Berg, 1999) on a log-log scale is quite close to a straight line.¹
This inverse linear (p.381)

(p.382)
relationship between the log of the magnitude of a billionaire's fortune and the person's logarithmic rank suggests that the wealth distribution in the *Forbes* list follows a power-law distribution (Levy & Solomon, 1997). Perhaps the most well-known instance of a power-law distribution in the social sciences is *Zipf's law*. In his book *Human Behavior and the Principle of Least Effort*, George Kingsley Zipf (1949) observed that rank-size distributions in domains as diverse as city sizes and word frequencies can

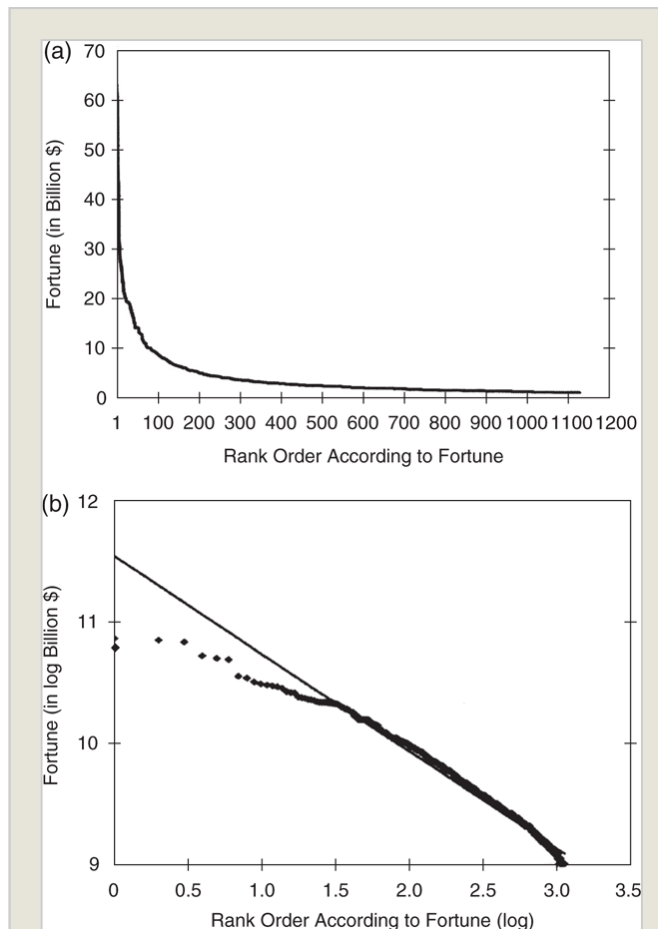


Figure 15-1: The world's 1,125 billionaires in 2008 rank ordered by fortune. (a) Absolute data. (b) Same data but with the logarithmic values (base 10) of ranks and fortune. The solid line corresponds to the least-square fit and has a slope of -0.78 . The approximate straight-line form implies that the distribution follows a power law (see also Levy & Solomon, 1997). Data from the *Forbes* magazine 2008 survey (Kroll, 2008).

How Estimation Can Benefit From an Imbalanced World

be described by a straight line in a log-log plot, whose slope q equals -1 . In the context of city sizes, this slope means that the population of a city is inversely proportional to its rank: Consequently, the second-ranked city in a country has half the population of the biggest city, the third-ranked city one-third that population, and so on. The rank-city size distributions for cities within one country appear to fit Zipf's law remarkably well.² In terms of a probability distribution, this means that the probability that the size of a city (or any other object) is greater than some S is proportional to $1/S$: $P(\text{Size} > S)$

S^q , with $q \approx -1$ (Gabaix, 1999).

Power-law distributions occur in an extraordinarily diverse range of domains, for instance, the sizes of earthquakes, firms, meteorites hitting the earth, moon craters, solar flares, and computer files; the intensity of wars; the frequency of use of words in any human language or of occurrence of personal names in most cultures; the numbers of papers that scientists write, of citations received by papers, of hits received by websites, of telephone calls made; the sales of books and music recordings; the number of species (p.383) in biological taxa; and the likelihood that a record in memory will be needed (see Bak, 1997; Buchanan, 1997; Krugman, 1996; Lehman, Jackson, & Lautrup, 2006; Newman, 2005; Schroeder, 1991).

Although Pareto's notion of "predictable imbalance" originally referred to income distributions, we use it here to describe the phenomenon of pronounced environmental skewness that is characteristic of power-law distributions: Few objects take on very large values (e.g., frequency, intensity, size) and most take on medium to small values. In high-energy physics, for instance, about half of all papers receive two or fewer citations, and the top 4.3% of papers produces 50% of all citations, whereas the bottom 50% of papers yields just 2.1% of all citations (Lehman et al., 2006). Income inequality is not just a phenomenon found in the exclusive circle of billionaires but also among street gangs. In one analysis of a Chicago street gang, the Black Disciples, the top 120 men—representing just 2.2% of the gang membership—took home

well more than half the money the gang accrued (Levitt & Dubner, 2005, p. 103). Environment imbalance is also ubiquitous in consumer markets. Take, for example, the success of Hollywood movies measured in terms of their box office gross. According to Anderson (2006), an estimated 13,000 feature films are shown in film festivals each year in the United States alone. They can be arranged into three groups. The first includes the 100 movies with the highest revenue, the ones that knocked out audiences. The second group of movies, those of rank 101 to 500, make low but not quite zero revenues, and the sorry remainder, rank 501 to 13,000, have no box office gross (mostly because they did not even garner mainstream commercial distribution). Anderson referred to such a distribution as “the Long Tail” (adapting the notion of long-tailed distributions from statistics), and he saw them everywhere in markets.

The question that concerns us here is this: Given that predictable imbalance is such a ubiquitous environmental structure, could it be that particular human cognitive strategies have evolved or been learned to exploit it?

QuickEst: A Fast and Frugal Estimation Heuristic in a World Full of Power-Law Regularities

Enrico Fermi, the world-renowned physicist and one of the leaders of the team of physicists on the Manhattan Project that eventually led to the development of the atomic bomb, had a talent for quick but reliable estimates of quantities. Legend has it that in the Alamogordo Desert in the state of New Mexico, while banks of spectrograph and ionization chambers waited to be triggered into action to assimilate the complex signals of the first atomic explosion, (p.384) Fermi was awaiting the same detonation from a few thousand yards away. As he sheltered behind a low blast-wall, he tore up sheets of paper into little pieces, which he tossed into the air when he saw the flash. After the shock wave passed, he paced off the distance traveled by the paper shreds, performed a quick back-of-the-envelope calculation, and arrived at an approximately accurate figure for the explosive yield of the bomb (Logan, 1996). For Fermi, one of the most important

How Estimation Can Benefit From an Imbalanced World

skills a physicist ought to have is the ability to quickly derive estimates of diverse quantities. He was so convinced of its importance that he used to challenge his students with problems requiring such estimates—the fabled canonical Fermi problem was the question: “How many piano tuners are there in Chicago?”

Being able to make a rough estimate quickly is important not only for solving odd Fermi problems. There is ample opportunity and need for people to rely on quick and easy estimates while navigating through daily life (e.g., how long will it take to get through this checkout line?). How do people arrive at quick quantitative estimates? For instance, how do they swiftly estimate the population size of Chicago—a likely first step toward an estimate of the number of piano tuners in Chicago? Previously, we have argued that cognitive estimation strategies, specifically, the *QuickEst heuristic*, may have evolved to exploit the predictable imbalance of real-world domains so as to reduce the computational effort and informational demands needed to come up with competitively accurate estimates (Hertwig, Hoffrage, & Martignon, 1999). In this chapter, we analyze the ecological rationality of this heuristic in more precise terms: First, we quantify the degree of imbalance across a total of 20 real-world domains using the parameter q , the slope of the straight line fitting the log-log rank-size distribution. Second, we analyze to what extent this degree of imbalance and other statistical properties of those environments hinder or foster the accuracy of the QuickEst heuristic. Before we turn to this analysis, we describe QuickEst in more detail.

The QuickEst heuristic is a model of quantitative *inferences from memory* (Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Goldstein, 2008), that is, inferences based on cue information retrieved from memory. It estimates quantities, such as the size of Chicago or the number of medals that Russia won at the most recent Olympic summer games. In general, it estimates the value of an item a , an element of a set of N alternatives (e.g., objects, people, events), on a quantitative criterion dimension (e.g., size, age, frequency). The heuristic’s estimates are based on M binary cues (1, 2, ...,

How Estimation Can Benefit From an Imbalanced World

i, \dots, M), where the cue values are coded such that 0 and 1 tend to indicate lower and higher criterion values, respectively. As an illustration, consider the reasoning of a job candidate who is subjected to a brainteaser interview by a company recruiter. (p.385) One task in the interview is to quickly estimate the net worth of, say, Donald Trump. To infer an answer the candidate may rely on cues such as: “Did the person make the fortune in the computer industry?”

To operate, QuickEst needs a set of cues put into an appropriate order. This order is based on the following measure: For any binary cue i , one can calculate the average size s_i^- of those objects that do not have the property that cue i represents. For instance, one can calculate the average net worth of all billionaires who are *not* entrepreneurs in the computer industry. The QuickEst heuristic assumes that cues are ranked according to the sizes of the values s_i^- , with the smallest value first.

In addition to the search rule, QuickEst also includes stopping and decision rules. The complete steps that the heuristic takes to estimate the criterion for object a are as follows:

Step 1: *Search rule*. Search through cues in the order of the sizes of the value s_i^- , starting with the smallest value.

Step 2: *Stopping rule*. If the object a has the value 0 on the current cue (indicating a low value on the criterion), stop searching and proceed to step 3. Otherwise (if the object has cue value 1 or the value is unknown), go back to step 1 and look up the cue with the next smallest s_i^- . If no cue is left, put the object into the catchall category.³

Step 3: *Decision rule*. Estimate the size of the object as the s_i^- of the cue i that stopped search, or of the size of the catchall category (see Hertwig et al., 1999, p. 225).

Estimates are finally rounded to the nearest spontaneous number.⁴

QuickEst’s structure maps onto the predictable imbalance of many real-world J-shaped environments (as in Figure 15-1). First, its asymmetric stopping rule—stop when a cue value of

zero is found for the object—limits search most strongly in environments in which zero (or absent) cue values are plentiful (cf. chapter 10). Second, by also first looking up the “small” cues—those cues (p.386) whose absence is associated with small criterion values s^- —QuickEst has an in-built bias to estimate any given object as relatively small. This is appropriate in the many J-shaped environments in which most objects have small values on the criterion, and only a few objects have (very) large values. Finally, QuickEst’s cue order also enables it to estimate small objects (with predominantly zero values on the cues) by looking up only one or a few (known) cues before providing an estimate—making it fast and frugal.

How Accurate Is QuickEst?

Can such a simple and fast estimation strategy nonetheless arrive at competitively accurate inferences? We compared QuickEst to two other estimation strategies, namely, *multiple regression* and an *estimation tree* that we designed (see Hertwig et al., 1999, for a detailed description of the estimation tree). Briefly characterized, multiple regression is a computationally powerful competitor insofar as it calculates weights that minimize least-squares error, and consequently it reflects the correlations between cues and criterion and the covariance between cues. The estimation tree arrives at estimates by collapsing objects, say cities, with the same cue profile (i.e., the same cue value on each of the available cues) into one class (for more on tree-based procedures, see Breiman, Friedman, Olshen, & Stone, 1993). The estimated size for each city equals the average size of all cities in that class (the estimate for a city with a unique cue profile is just its actual size). When the tree encounters a new, previously unseen city whose cue profile matches that of one or more previously seen cities, its estimated size is the average size of those cities. If a new city has an entirely new cue profile, then this profile is matched to the profile most similar to it. The estimation tree is an exemplar-based model that keeps track of all exemplars presented during learning as well as their cue values and sizes. As long as the test set and training set are identical, this algorithm is optimal. Yet, when the training set

How Estimation Can Benefit From an Imbalanced World

is large, it requires vast memory resources (for the pros and cons of exemplar-based models, see Nosofsky, Palmeri, & McKinley, 1994).

All three strategies were tested in the environment of 82 German cities with more than 100,000 residents (excluding Berlin). The task was to predict the cities' number of residents. This demographic target criterion follows a power law, thus exhibiting the property of predictable imbalance (remember that city size distributions were one of the classic domains in which Zipf, 1949, observed his law). To examine the strategies' *robustness*, that is, their ability to predict new data (here, cities), Hertwig et al. (1999) distinguished between two sets of objects: the *training* set and the *test* set. The strategies learned their parameters (e.g., s_i^- or beta weights) on (p.387) the basis of the training set. The test set, in turn, provided the test bed for the strategies' robustness. The training samples consisted of 10%, 20%, ..., 90%, and 100% of the 82 cities, comprising their population sizes and their values on eight cues indicative of population size. The test set encompassed the complete environment of 82 cities. That is, the test set included all cities in the respective training set, thereby providing an even harder test for QuickEst, because parameter-fitting models like multiple regression are likely to do relatively better when tested on objects they were fitted to.

In the environment of German cities, QuickEst, on average, considered only 2.3 cues per estimate as opposed to 7.3 cues used by multiple regression and 7.1 (out of 8) used by the estimation tree. Despite relying on only about a third of the cues used by the other strategies, QuickEst nonetheless exceeded the performance of multiple regression and the estimation tree when the strategies had to rely on quite limited knowledge, with training sets ranging between 10% and 40%. The 10% training set exemplified the most pronounced scarcity of information. Faced with such dire conditions, QuickEst's estimates in the test set were off by an average of about 132,000 inhabitants, about half the size of the average German city in the constructed environment. Multiple regression and the estimation tree, in contrast, erred

How Estimation Can Benefit From an Imbalanced World

on average by about 303,000 and 153,000 inhabitants, respectively.

When 50% or more of the cities were first learned by the strategies, multiple regression began to outperform QuickEst. The edge in performance, however, was small. To illustrate, when all cities were known, the estimation errors of multiple regression and QuickEst were 93,000 and 103,000 respectively, whereas the estimation tree did considerably better (65,000).⁵ Based on these results, Hertwig et al. (1999) concluded that QuickEst is a psychologically plausible estimation heuristic, achieving a high level of performance under the realistic circumstances of limited learning and cue use.

How Robust Is QuickEst's Performance Across Diverse Environments?

Although QuickEst competitively predicted demographic quantities, we did not know how well its competitiveness would generalize to other environments—in particular, to environments that exhibit (p.388) different degrees of predictable imbalance. Our first goal in this chapter is to investigate this issue. To this end, we test QuickEst, multiple regression, and the estimation tree with a collection of 20 different real-world environments. As previously, we take from each environment increasingly larger portions from which the strategies can learn. This emphasis on learning reflects the typical situation of human decision making, an issue to which we return shortly. Again, the training sets consist of 10%, 20%, ..., 90%, and 100% of each environment. To arrive at psychologically plausible sets of limited object knowledge, we also assume that the probability that an object belongs to the training set is proportional to its size (thus capturing the fact that people are more likely to know about larger objects than smaller ones). The predictive accuracy of the strategies is tested on the complete environment (i.e., the test set; as in Hertwig et al., 1999, the training set is a subset of the test set). To obtain reliable results, 1,000 random samples are drawn for 9 of the 10 sizes of the training set (in the 100% set, training set equals test set, and thus sampling error is of no concern).

How Estimation Can Benefit From an Imbalanced World

For the environments, we make use of the collection of real-world data sets that Czerlinski, Gigerenzer, and Goldstein (1999) compiled to test the performance of fast and frugal choice strategies. This collection includes such disparate domains as the number of car accidents on a stretch of highway, the homelessness rate in U.S. cities, and the dropout rates of Chicago public high schools. The environments ranged in size from 11 objects (ozone levels in San Francisco measured on 11 occasions) to 395 objects (fertility of 395 fish), and included 3 to 18 cues. All cues were binary or were made binary by dichotomizing them at the median. One particularly attractive aspect of this collection of environments is that Czerlinski et al. did not select them to match any specific distribution of the criterion, with many of these environments taken from textbook examples of the application of multiple regression. On average, these environments were not as skewed as, for instance, the myriad real-world environments from which Zipf (1949) derived his eponymous law. The median q in this set of environments is -0.54 , and thus substantially smaller in magnitude than the $q \approx -1$ that Zipf observed (see also Newman, 2005, who found a median exponent of -2.25 in his broad set of distributions of quantities measured in physical, biological, technological, and social systems).

How Frugal Are the Strategies?

QuickEst is designed to make estimates quickly, using few cues. This ability became manifest in the present simulations. Figure 15-2 shows the number of cues that QuickEst considered as a function (p.389)

How Estimation Can Benefit From an Imbalanced World

of the size of the training set. Across all environments, 7.7 cues, on average, are available. QuickEst considers, on average, only two cues (i.e., 26%) per estimate—a figure that remains relatively stable across various sizes of training set size. In contrast, multiple regression

(which here uses only those cues whose beta weights are significantly different from zero) and the estimation tree use more and more cues with increasing training sets. Across all training set sizes, they use an average of 5.1 (67%) and 5.9 (77%) of all available cues, respectively.

How Robust Are the Strategies?

What price does QuickEst pay for betting on J-shaped environment structures, and for considering substantially fewer cues than its competitor strategies? The first benchmark we use to answer this question is *robustness*. Robustness describes the strategies' ability to generalize from small training sets to the test set. We first calculate the strategies' absolute errors (i.e., absolute deviation between actual and estimated size) separately for each environment and training set. Then, we define each strategy's performance in the 100% training set as the strategy's maximum performance and express the absolute errors observed in all other training sets as a percentage of this maximum-performance benchmark (e.g., if a strategy makes errors of 60,000 with the 100% training set and 90,000 with the 40% training set, then for the

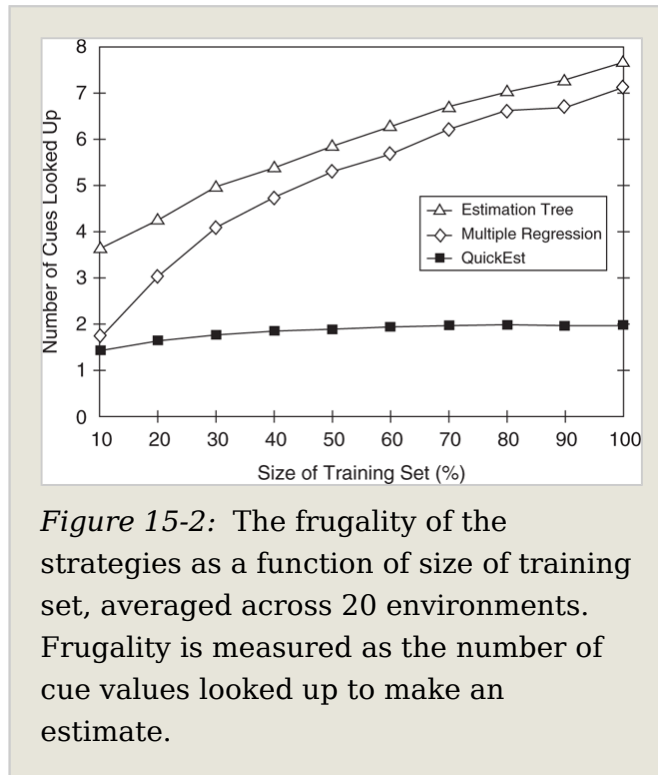


Figure 15-2: The frugality of the strategies as a function of size of training set, averaged across 20 environments. Frugality is measured as the number of cue values looked up to make an estimate.

How Estimation Can Benefit From an Imbalanced World

latter it would have (p.390) a normalized error of 150%).

Finally, we average these normalized estimation errors (which must by definition be above 100%) across all environments, separately for each strategy and each training set size.

Based on this mean, we can define robustness as the resistance to relative decline in performance as training sets become smaller. Figure 15-3 shows the normalized estimation error (averaged across the 20 environments). QuickEst proves to be a robust strategy. When only 40% of the environments' objects are learned, QuickEst still performs about as well as when all objects are known. Moreover, when QuickEst is required to rely on a very thin slice of the environments, as exemplified by the 10% training set, its error is only about 1.5 times the magnitude of its maximum-performance error. Multiple regression and the estimation tree, in contrast, are less robust. When 50% of the objects are known, for example, their respective errors are about 1.5 and 3 times the magnitude of their maximum-performance error. Their relative lack of robustness becomes most pronounced under extreme scarcity of information. In the 10% training set, their error is more than 2 times (multiple regression) and 6 times (estimation tree) the size of their maximum-performance errors.

In generalizing to unknown territory, QuickEst thus suffers less than do some computationally and informationally more expensive

How Estimation Can Benefit From an Imbalanced World

(p.391)

strategies. The ability to generalize to new data appears to be a key property of efficient human decision making. In most real-world environments people cannot help but act on the basis of scarce knowledge.⁶ In fact, scarcity of knowledge is a crucial human condition, as is suggested by, for instance, Landauer's

(1986) analysis of how much information is accumulated in a single human's memory over the course of a normal lifetime. Basing his calculations on various bold assumptions (e.g., about the rate at which people can take in information), he estimated that the "functional learned memory content" is "around a billion bits for a mature person" (p. 491). In comparison, an institutional memory of human knowledge such as the Library of Congress with 17 million books is estimated to contain about 136 terabytes—about 1,088 trillion bits, more than one million times the estimated magnitude of human memory (Lyman & Varian, 2003). Although Landauer's figure is an audacious (if scientifically informed) estimate, it supports the notion that most of human decision making occurs under conditions of scarcity of information and knowledge. Upon these terms, frugality and robustness appear to be key properties of competitive cognitive strategies.

How Accurate Are the Strategies?

Although the previous analysis demonstrates QuickEst's robustness, measured in terms of how little its performance

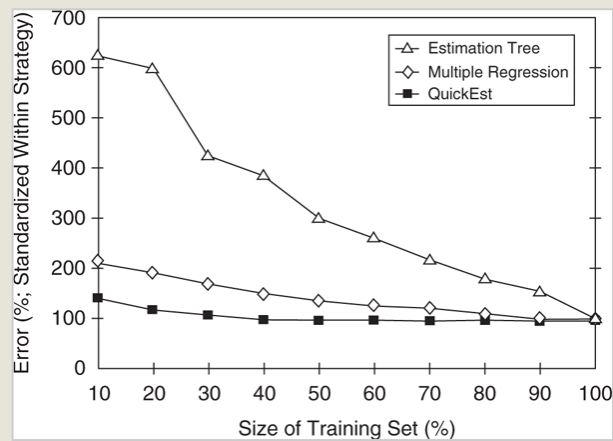


Figure 15-3: The estimation error (standardized within each strategy) as a function of size of training set, averaged across 20 environments. For each strategy, we standardized its accuracy by expressing its error per training set relative to its estimation error made in the 100% training set (i.e., the error of each strategy under complete knowledge was assumed to be 100%).

How Estimation Can Benefit From an Imbalanced World

deteriorates with smaller and smaller training sets, it says nothing about the heuristic's accuracy relative to its competitors. In fact, if we equate needing less information with involving less effort, the well-known effort-accuracy tradeoff (Payne, Bettman, & Johnson, 1993) would predict that this decreased effort goes along with decreased accuracy. So does QuickEst's robustness come at the price of lower accuracy compared to its more effortful competitors? To test for this possibility, we next compare QuickEst's estimation accuracy with that of its rivals. To this end, we now treat QuickEst's maximum performance (with the 100% training set) as the benchmark and express its own performance and that of its competitors relative to this benchmark set at 100%. Figure 15-4 shows the (p.392)

strategies' relative estimation error as a function of the training set size (the line for QuickEst being the same as in Figure 15-3).

Several results are noteworthy: QuickEst's performance under scarcity of knowledge is not inferior to that of its competitors. On the contrary, it is here that QuickEst outperforms the other strategies. In the 10% training set, for instance,

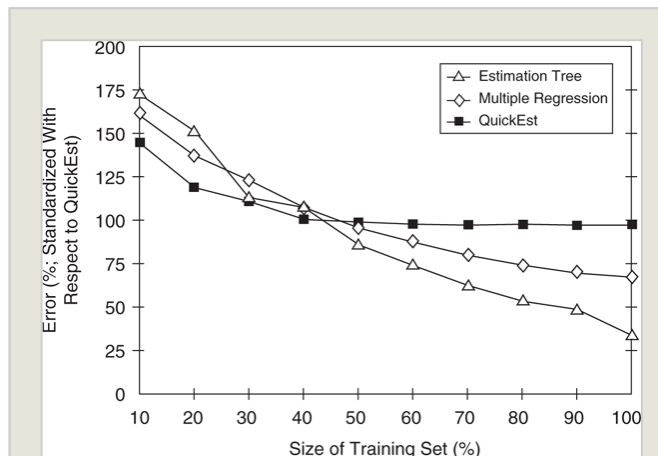


Figure 15-4: The estimation error (standardized with respect to QuickEst's performance) as a function of size of training set, measured across 20 environments. For each strategy, we standardized its accuracy by expressing its error per training set relative to QuickEst's estimation error made in the 100% training set (i.e., QuickEst's error under complete knowledge was assigned to be 100%).

QuickEst's error amounts to 1.45 times the size of the error it produced with the 100% training set. In contrast, errors with multiple regression and the estimation tree in the 10% training set are 1.6 and 1.7 times higher than for the 100% training set, respectively. Moreover, as long as the training set encompasses less than 50% of the environment, QuickEst either outperforms its competitors or matches their performance. Only when the training set is 50% and larger does QuickEst fall behind. In fact, under the circumstances of complete knowledge (100% training set), QuickEst is clearly behind multiple regression and the estimation tree: The magnitude of their error is about 0.7 and 0.4 times the size of QuickEst's error, respectively.

In sum, QuickEst outperforms multiple regression and the estimation tree when knowledge is scarce. In the psychologically (p.393) less plausible situation of abundant knowledge (i.e., 50% or more of the environments' objects are known) QuickEst, however, clearly falls behind the performance of its competitors. All these results are based on the strategies' performance averaged across 20 quite different environments. Now, we turn to our next question: Which statistical properties of the environments predict differences in performance between QuickEst and the other strategies?

Which Environment Properties Determine QuickEst's Performance?

We focus on three important properties of environments: variability, skewness, and object-to-cue ratio (see chapter 4 for a discussion of the first two). Variability refers to how greatly the objects in an environment vary from the mean value of that set of data. We quantify this property by calculating each environment's *coefficient of variation* (CV):

$$CV = \frac{SD}{\text{mean}}$$

which is the ratio of the standard deviation (*SD*) of the set of object criterion values to its mean value.

The next property, skewness, captures how asymmetric or imbalanced a distribution is, for instance, how much of a "tail" it has to one side or the other. Skewness can be measured in

How Estimation Can Benefit From an Imbalanced World

terms of the parameter q , estimated with the following method (Levy & Solomon, 1997): We sort and rank the objects in each environment according to their criterion values, and fit a straight line to each rank-size distribution (plotted on log-log axes). We then use the slope q of this fitted regression line as an estimate of the environment's skewness.

The final property in our analysis is the object-to-cue ratio (i.e., the ratio between the number of objects and number of cues in an environment), which has been found to be important in the analysis of inferential heuristics such as take-the-best (see Czerlinski et al., 1999; Hogarth & Karelaia, 2005a). To assess the relationship between the statistical properties of the environments and the differences in the strategies' performance, we first describe the results regarding skewness for two environments in detail, before considering all 20 environments.

Two Distinct Environments: U.S. Fuel Consumption and Oxygen in Dairy Waste

Does an environment that exhibits predictable imbalance, or skew, such that few objects have large criterion values and most (p.394) objects take on small to medium values, foster the performance of QuickEst? And, vice versa, does a more balanced, that is, less skewed environment impair QuickEst's performance? The most imbalanced environment in our set of 20 is the *oxygen* environment ($q = -1.69$; with a fit of the regression line of $R^2 = .98$). Here, the task is to predict the amount of oxygen absorbed by dairy wastes from cues such as the oxygen required by aerobic micro-organisms to decompose organic matter. The *fuel consumption* environment, in contrast, is relatively balanced, with a q parameter that is about eight times smaller ($q = -0.2$; $R^2 = .87$). Here, the task is to predict the average motor fuel consumption per person for each of the 48 contiguous U.S. states from cues such as state fuel tax and per capita income. The environments' markedly different degree of imbalance is illustrated in Figure 15-5. The rank-size distributions (in logarithmic scales) yield the characteristic negative-sloping linear relationship, thus suggesting that the power law provides a good model for both environments.

How Estimation Can Benefit From an Imbalanced World

Is the difference in environmental skewness predictive of the strategies' performance? Figure 15-6 shows the strategies' relative error as a function of the training set and the two environments. Figure 15-6a plots the results for the highly skewed oxygen environment. QuickEst's performance is strongly competitive: Across all training set sizes, QuickEst consistently outperforms multiple

(p.395)

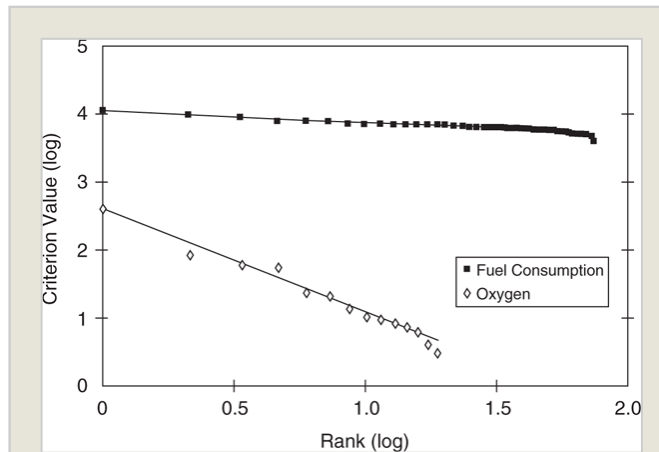


Figure 15-5: Log-log scale plot of the distribution of dairy wastes rank ordered by their amount of oxygen absorbed (oxygen), and the distribution of 48 U.S. states rank ordered by their average motor fuel consumption (fuel consumption). Each plot also shows the regression line fitted to the data.

How Estimation Can Benefit From an Imbalanced World

(p.396)

regression. In addition, the estimation tree can only outperform QuickEst (and by a small margin) when it learns about 70% or more of the objects in the environment. Finally, under the psychologically unlikely circumstance of complete knowledge (100% training set), QuickEst's performance is only six percentage points below the estimation tree's performance. The picture looks strikingly different in the far less imbalanced fuel consumption environment (Figure 15-6b). Except for the 10% training set, multiple regression and the estimation tree consistently outperform QuickEst. This contrast between the two environments suggests that QuickEst's performance, relative to that of its competitors, hinges on environmental skewness. We shall now see to what extent this observation generalizes across all environments.

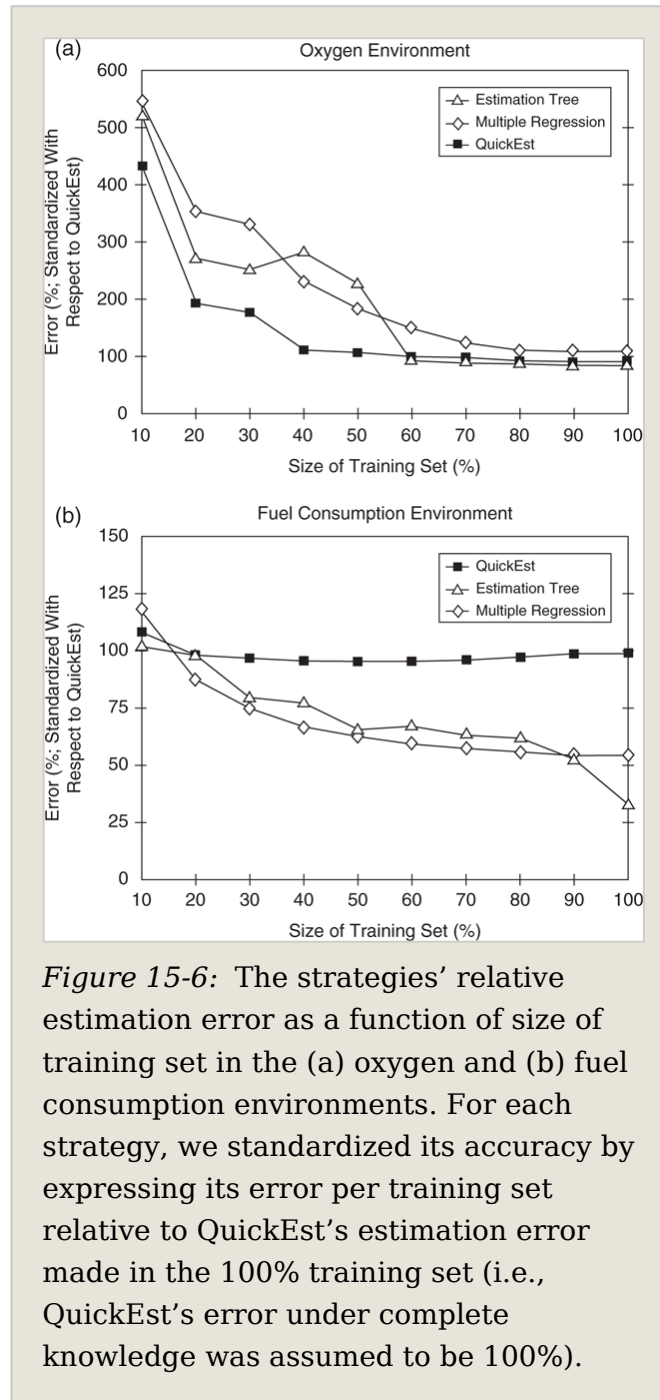


Figure 15-6: The strategies' relative estimation error as a function of size of training set in the (a) oxygen and (b) fuel consumption environments. For each strategy, we standardized its accuracy by expressing its error per training set relative to QuickEst's estimation error made in the 100% training set (i.e., QuickEst's error under complete knowledge was assumed to be 100%).

How Estimation Can Benefit From an Imbalanced World

Can Environmental Skewness and Variability Explain QuickEst's Failures and Successes?

The environmental parameter q is a measure of the amount of skewness in the criterion distribution: The smaller q is, the flatter the distribution, and vice versa. In our set of 20 environments, skewness varies widely, ranging from -0.02 to -1.69 , with a median of -0.54 . Does greater skewness in the criterion distribution contribute to better QuickEst performance, relative to its competitors?

Figure 15-7 shows that QuickEst's performance indeed depends on the environments' skewness: Its advantage over multiple regression (measured in terms of QuickEst's relative error minus multiple regression's relative error) is most pronounced in environments with large (negative) q . Relatedly, multiple regression tends to outperform QuickEst in environments with small q . The correlation between the difference in the strategies' errors and the magnitude of q is .86. For illustration, the largest magnitudes of q and hence greatest skewness occur in the oxygen ($q = -1.69$), biodiversity ($q = -1.6$), and mammals' sleep environments ($q = -1.14$). It is in these environments that the largest advantage of QuickEst over multiple regression can also be observed. In contrast, the largest advantages of multiple regression over QuickEst coincide with q values that are an order of magnitude smaller than those observed in the most skewed environments (obesity environment: $q = -0.08$; body fat environment: $q = -0.02$). This pattern also generalizes to the comparison of QuickEst and the estimation tree (not shown): Here, the correlation between the difference in the strategies' relative errors and q amounts to .8.

Environmental skewness implies variability in the criterion distribution, but variability does not necessarily imply skewness. Therefore, variability, independent of skewness, may be predictive of QuickEst's performance. In our set of environments, the coefficient (p.397)

How Estimation Can Benefit From an Imbalanced World

of variation varies widely, ranging from the oxygen environment, in which the standard deviation is twice as large as the mean ($CV = 2$), to the body fat environment, in which the standard deviation is a tiny fraction of the mean ($CV = 0.019$). We found that QuickEst has a clear advantage over multiple regression in environments with high variance (with

advantage again measured in terms of the difference between QuickEst's relative error and that of multiple regression). Across all environments, the correlation between the difference in the two strategies' relative errors and the CV is .87 (for the comparison with the estimation tree the correlation amounts to .8). In the current collection of environments, however, CV does not explain more regarding QuickEst's performance than does environmental skewness. This is not too surprising given that across environments, the Pearson correlation between parameter q and the coefficient of variation is $-.96$.

Is the Ratio of Objects to Cues Indicative of QuickEst's Performance?

When multiple regression is used as a strategy to model choice between two objects, it typically estimates first the criterion value (e.g., salary) separately for each object and then compares the objects. (p.398) Thus used, estimation is a

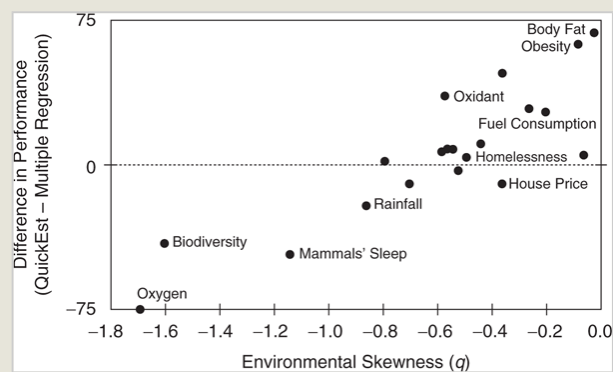


Figure 15-7: QuickEst's performance relative to multiple regression (in terms of the relative estimation error for QuickEst minus that for multiple regression; see Figure 15-4), plotted against skewness parameter q (the slope of the straight line fitted into the rank-size distributions of the current collection of environments) for the 20 environments. Negative values on the y-axis indicate an advantage of QuickEst over multiple regression; positive values indicate a disadvantage.

How Estimation Can Benefit From an Imbalanced World

precursor to choices. In the context of choices, in turn, it has been shown that multiple regression can be outperformed by simpler strategies (with unit weights) when the ratio between objects and cues becomes too small (Dawes, 1979; Einhorn & Hogarth, 1975; Schmidt, 1971; see also chapter 3). A statistician's rule of thumb is that unit weights will outperform regression weights if the latter are based on fewer than 10 objects per cue. The reason is that multiple regression is likely to grossly overfit the data when there are too few objects for the number of cues (see also Czerlinski et al., 1999).

Is the object-to-cue ratio also indicative of performance in the present context in which the task is to estimate the quantitative value of an individual object? Across the 20 environments, there is no substantial correlation (.08) between the object-to-cue ratio and the difference in relative errors between multiple regression and QuickEst. The correlation, however, increases (to .42) if one excludes the fish fertility environment, in which the object-to-cue ratio is extreme with 395 objects and three cues. This higher correlation suggests that QuickEst (like unit-weight decision heuristics) tends to have an advantage over multiple regression when there are fewer objects per cue.⁷ Yet, compared with the impact of skewness and variance, the object-to-cue ratio is a mediocre predictor of QuickEst's performance.

In sum, we examined several properties of ecological structures and found one that proved outstanding in its ability to predict QuickEst's performance (see also von Helversen & Rieskamp, 2008): The more skewed (and in the set we evaluated, the more variable) an environment, the better QuickEst performs in relation to its competitors. The correlation between the skewness q and the performance of QuickEst relative to that of multiple regression was .86; the correlation for QuickEst relative to the estimation tree was .8.

How Can People Tell When to Use QuickEst?

A heuristic is not good or bad, not rational or irrational, in itself, but only relative to an environment. Heuristics can exploit regularities in the world, yielding ecological rationality.

How Estimation Can Benefit From an Imbalanced World

QuickEst wagers that the criterion dimension is distributed such that few objects are very large, and most objects are relatively small (Hertwig et al., 1999). If QuickEst's wager on the environment structure matches the actual structure of the environment, it can perform (p.399) well. If QuickEst mismatches the environment structure, it will have to foot the bill for its bet.

Looking at the characteristics of particular environments in which the different estimation strategies excel, we found that QuickEst outperforms—even under conditions of abundant knowledge—multiple regression and estimation trees in environments with pronounced skewness and variability: The more skewed and variable the criterion value distribution in an environment, the better QuickEst's performance was relative to its competitors.

Given their fit to particular environment structures, using fast and frugal heuristics successfully means using them in the proper domains. But how can people tell what is a proper domain for a particular strategy, and what is improper? We suggest that the task of strategy selection may not be as taxing as it is often conceived. Let us distinguish between two kinds of "proper" environments. One is the class of environments in which people can muster little to medium knowledge. As the current simulations and those involving other fast and frugal strategies (Gigerenzer, Czerlinski, & Martignon, 1999) have shown time and again, the more limited the knowledge about an environment is, the more competitive simple strategies are. Their simplicity renders the heuristics robust and successful relative to more complex information-demanding strategies—even if the heuristics' match to the environment is not perfect.

A second class of "proper" environments is one in which users of, for instance, QuickEst can intuit that the structure of the environment maps onto the structure of the heuristic. To be able to do so, however, does not mean that people need to fit a power-law model to their knowledge, thus estimating the skewness of the environment. There are simple shortcuts instead that can gauge skewness. For instance, in environments with a very pronounced level of predictable

How Estimation Can Benefit From an Imbalanced World

imbalance, most objects one knows will have criterion values below the average (see the example of above-average drivers in chapter 4). Thus we propose that a mean value that substantially exceeds the median value may trigger the use of QuickEst. For instance, if a decision maker applied QuickEst in only those environments in which the mean value is, say, at least 50% greater than the magnitude of the median value, then in the current collection of 20 environments (and averaged across all training sets), QuickEst would be employed in four environments. In all of those QuickEst outperforms multiple regression, whereas multiple regression outperforms QuickEst in 13 of the remaining 16 environments. Thus, the ratio mean-to-median is a good proxy for the relative performance of the two strategies. This is consistent with our previous analysis, according to which skewness and the coefficient of variation proved to be good predictors of QuickEst's relative (p.400) performance—the ratio of mean-to-median correlates highly with both environmental properties ($-.81$ and $.92$, respectively).

On the basis of these two classes of “proper” environments, one can also deduce a class of environments that is “improper” for simple heuristics. It encompasses those environments in which people possess much knowledge *and* in which the structure of the heuristic mismatches that of the environment (e.g., for QuickEst this would mean that there is little skew in the distribution of criterion values). But the chance of erroneously applying a fast and frugal strategy like QuickEst in such improper environments may be slim, because having abundant knowledge should make it more likely that people have a sense of the environment's structure. However, do people always rely on QuickEst if the environment is skewed? And what strategies are used in environments that are not skewed? Next, we introduce another tool of the adaptive toolbox, the mapping heuristic (von Helversen & Rieskamp, 2008), which can be successfully employed in environments with different types of structure.

The Mapping Heuristic: A Tallying Approach to Estimation

Like QuickEst, the mapping heuristic is a simple strategy for making quantitative estimations from multiple cues, and it, too, relies on binary cue information.⁸ The estimation process is split into a categorization phase and an estimation phase. First, an object is categorized by counting all the positive cue values it has. Then, the mapping heuristic estimates the object's size to be the typical (median) size of all previously seen objects in its category, that is, with the same number of positive cues. This estimation strategy implies that all cues are treated as being equally important. Thus, in contrast to QuickEst, which considers cues sequentially, the mapping heuristic takes a tallying approach. It includes all relevant cues but weights each cue the same, ignoring the different predictive values of the cues. The two heuristics represent different approaches to simplifying the estimation process—ordered and limited cue search (see chapter 10) versus equal-weight tallying of all cues. How do the two approaches compare in terms of their performance in different environments?

To test when QuickEst and the mapping heuristic perform well and how much their performance depends on the structure of the environment (in terms of the distribution of the criterion), von p.401 Helversen and Rieskamp (2008) conducted a simulation study. Two types of environment were used, one with a skewed criterion (based on a power function $y = bx^a$, with $a = -1$, $b = 100$) and one involving a uniformly distributed criterion (based on a linear function, $y = bx + c$, with $b = -2$ and $c = 102$). For each distribution, several instances of the corresponding environments were generated, systematically varying the average correlation of the cues with the criterion and the number of positive cue values. Each environment consisted of 50 objects and five binary cues.

In addition to evaluating QuickEst and the mapping heuristic, the simulations also compared the estimation performance of multiple linear regression and an exemplar-based model (Juslin, Olsson, & Olsson, 2003) similar to the estimation tree.

How Estimation Can Benefit From an Imbalanced World

The accuracy of the models was determined by using a split-half cross-validation procedure, with each data set split 100 times in two halves. The models were fitted to the first half, the training set, to determine the values of the models' parameters. With these parameters the models made predictions for the second half of the data, the test set. The accuracy of these predictions was evaluated by determining the root mean square deviation (RMSD) between them and the actual criterion values, averaged separately across all skewed and uniform environments.

As expected, the more complex models, multiple linear regression and the exemplar model, achieved a better fit than the simpler QuickEst and the mapping heuristic on the training sets in both types of environments (Table 15-1). However, when generalizing to predictions in the test set, both heuristics outperformed the complex models. Von Helversen and Rieskamp found that, consistent with the results of the simulations reported earlier in this chapter, QuickEst predicted best in the skewed environments, whereas

How Estimation Can Benefit From an Imbalanced World

Table 15-1: Average Model Accuracy (RMSD) for Different Environment Structures (as Criterion Distributions)

Model	Environment				Uniform			
	Skewed							
	Training set		Test set		Training set		Test set	
	M	SD	M	SD	M	SD	M	SD
QuickEst	14.8	1.7	14.9	1.1	24.8	3.5	28.3	3.5
Mapping	14.3	3.5	15.3	1.6	21.6	5.1	25.9	6.4
Regression	14.0	2.4	16.5	1.2	20.9	4.7	27.7	6.3
Exemplar	12.0	3.5	15.8	1.7	17.5	4.9	27.2	6.2

Note. Lower values denote better performance.

(p.402) the mapping heuristic predicted best when the criterion was uniformly distributed. In addition, the mapping heuristic performed better than the regression model in both types of environments and thus was less dependent on the distribution of the criterion than QuickEst.

Which Strategy to Select From the Adaptive Toolbox?

When should people use QuickEst or the mapping heuristic? Which heuristic people apply should depend on the characteristics of the environment they are facing. This suggests that QuickEst should be chosen in skewed criterion distributions and the mapping heuristic should be recruited in uniform or less skewed distributions. In addition, we would like to introduce a second environmental structure that could influence the choice between QuickEst and the mapping heuristic: the dispersion of the cues. For inference strategies, it has been shown that a lexicographic heuristic like take-the-best, for instance, performs especially well when the cues have diverse validities and when the intercorrelations between the cues are high. In contrast, in situations with equally valid cues and low intercorrelation, a tallying heuristic that integrates the information of all available cues performs well (Dieckmann & Rieskamp, 2007; Hogarth & Karelaia, 2007; Martignon & Hoffrage, 2002; see also chapters 3, 8, and 13). Analogously, the cognitive processes that take place when people make estimations may depend on environmental features similar to those used in the selection of take-the-best or tallying. Thus, QuickEst could be particularly suited for skewed distributions with highly dispersed cue validities, whereas the mapping heuristic might be most suited when the cues have similar validities.

Do People Use Heuristics for Estimation?

Given these predictions about when each estimation strategy should be used to achieve ecological rationality, we can next ask whether people actually *do* use QuickEst and the mapping heuristic in particular appropriate environments. First, three recent experiments have looked at how well QuickEst describes the memory-based estimates that people make (as

How Estimation Can Benefit From an Imbalanced World

opposed to inferences from givens⁹). Woike, Hertwig, and Hoffrage (2009) asked people to (p.403) estimate the population sizes of all 54 countries in Africa and, in addition, probed their knowledge of numerous cues and cue values indicative of population size (e.g., membership in the Organization of the Petroleum Exporting Countries, location in the Sahel zone, etc.). People's actual estimates of the countries' population sizes were then compared to predictions from three distinct strategies, made using each individual's often very limited cue knowledge. The strategies were QuickEst, multiple regression, and *Probex*, an exemplar-based strategy that has been found to successfully model people's estimates of quantities such as city sizes (Juslin & Persson, 2002). The psychological models, QuickEst and *Probex*, both predicted people's estimates better than the statistical model, multiple regression. More specifically, QuickEst better predicted actual estimates of about three-fourths of the participants, whereas *Probex* proved to be the better model for the remaining quarter. In their second study using the same methodology, Woike et al. (2009) asked participants to estimate either African countries' population size (a J-shaped distribution) or their respective rate of illiteracy (a uniform distribution). In addition, participants indicated their knowledge of six cues related to either population size or illiteracy rate. As expected, QuickEst fared better than *Probex* in capturing people's estimates in the J-shaped environment, whereas *Probex* scored better in the uniform environment.

In another experiment asking participants to estimate city population sizes, Hausmann, Läge, Pohl, and Bröder (2007, Experiment 1) found no correlation between how long people took to arrive at an estimate of the size of a city and its estimated size. They took this to be evidence against the use of QuickEst, which they conjectured would predict a positive correlation because the heuristic's cue search should stop earlier for smaller than for larger cities. The correlation between size of cities and response time, however, is likely to be moderated by at least one factor, the retrieval speed of cue values. In fact, using a set of 20 German cities, Gaissmaier (2008) analyzed the retrieval speed of cue values as a function of city size. He found that the larger a city, the faster the

retrieval of its cue values (regardless of whether the cues indicated absence or presence of a property), and that it takes longer to retrieve the absence of a property (e.g., has no airport) for a small city than to (p.404) retrieve the presence of a property (e.g., has an airport) for a large city. These links between retrieval speed of cue values and size of objects can be understood within Anderson's ACT-R framework (Adaptive Control of Thought-Rational—see Anderson & Lebiere, 1998; Hertwig, Herzog, Schooler, & Reimer, 2008; see also chapter 6). Based on these observations, one can predict that the time one saves from the heuristic's frugality for small cities may be consumed by the longer retrieval times of small cities' cue values, relative to those for large cities. Counterintuitively—but consistent with the data of Hausmann et al.—QuickEst may therefore take equally long to arrive at estimates for small and large cities.

Two other experiments looked at how well the mapping heuristic predicted people's estimates (von Helversen & Rieskamp, 2008). These experiments involved inferences from givens rather than from memory, and participants used the given cues to make estimates in a task with either a skewed or a uniform criterion distribution. The mapping heuristic's prediction ability was then compared with two other estimation strategies: multiple regression and an exemplar-based model similar to Probox (Juslin et al., 2003). In both criterion distributions, von Helversen and Rieskamp found that the mapping heuristic, on average, predicted the estimates as well as or better than its two competitor models. Thus, the experimental evidence so far indicates that in both situations of inference from memory and inference from givens, simple fast and frugal mechanisms—whether QuickEst or the mapping heuristic—are often better at accounting for the estimates that people make than are more complex strategies.

How Does Predictable Environment Imbalance Emerge?

We used Pareto's notion of "predictable imbalance" to refer to the ubiquitous phenomenon of environmental skewness characteristic of power-law distributions: In many domains,

How Estimation Can Benefit From an Imbalanced World

few objects take on very large values (e.g., in frequency, intensity, size) and most take on medium to small values. What is the origin of such distributions? This is a hotly debated question, and the explanations of how such power-law distributions might arise in natural and man-made systems range from domain-general explanations such as “self-organized criticality” (e.g., Bak, 1997) to domain-specific explanations such as models of urban growth (e.g., Simon, 1955b) or the reasons for the rarity of large fierce animals (Colinvaux, 1978; see Newman, 2005, for a review of various explanations). In what follows, we briefly describe these two domain-specific accounts of predictable imbalance.

(p.405) Simon’s (1955b) model of urban growth aims to explain why rank-size distributions of city populations are often but not always nicely approximated by a straight line with a slope $q = -1$ (for examples see Brakman et al., 1999). It is assumed that new migrants to and from cities of particular regions arrive during each time period, and with a probability π they will form a new city, and with a probability of $1-\pi$ they will settle in a city that already exists (for an exposition of Simon’s model, see Krugman, 1996). The probability with which any given city attracts new residents is proportional to its size. If so, this model will generate a power law, with exponent $q = -1/(1-\pi)$, as long as π is very close to 0. In other words, if new migrants almost always join existing cities, then q will converge toward -1 . This elegant explanation of Zipf’s law for city-size distribution has, however, a number of drawbacks that various authors have pointed out (e.g., Krugman, 1996; Brakman et al., 1999).

In his book *Why Big Fierce Animals Are Rare*, the ecologist Paul Colinvaux (1978) concluded that body mass and metabolic demands of large animals set limits to their frequency. Indeed, as Carbone and Gittleman (2002) have shown, the relationship between the number of carnivores per 10,000 kg of prey and carnivore body mass itself follows a power function, with an exponent of -1 . For illustration, 10,000 kg of prey biomass cannot even support in perpetuity one polar bear whose average body mass amounts to 310 kg, whereas it supports 146 Channel Island foxes, which have an

average mass of about 2 kg. An adult male killer whale, with a daily caloric demand of 287,331 calories, must guzzle down five male or seven female sea otters per day, thus a single pod of killer whales (composed of one male and four females) could ingest over 8,500 sea otters per year (Williams, Estes, Doak, & Springer, 2004). Clearly, high caloric demands require a large intake of prey, and the question of why big fierce animals are rare comes down to whether these animals can find as much food as they need to survive.

Both domain-specific and domain-general scientific explanations have been proposed for ubiquitous types of statistical distributions, whether they be, for instance, power-law or Gaussian distributions. Assuming the human mind contains an adaptive toolbox of simple cognitive strategies (Gigerenzer, Czerlinski, et al., 1999), one unexplored issue is whether people have intuitive theories about the emergence of specific distributions—for example, “there need to be many, many more small animals than big animals, because any big one preys on many small ones”—and to what extent such theories play a role in triggering cognitive strategies that bet on specific types of distributions.

(p.406) Conclusion

Power-law distributions face us from all sides. Chater and Brown (1999) pointed out their ubiquity in environmental features that we perceive. Based on this, they argued that many psychological laws governing perception and action across domains and species (e.g., Weber’s law, Stevens’s law) reflect accommodation of the perceptuo-motor system to the skewed world. The same type of relationship to J-shaped environments has also been argued for the structure of memory (Anderson & Schooler; 1991; Schooler & Hertwig, 2005; see also chapter 6). Similarly, we take as a starting point the observation that power-law regularities hold across a wide range of physical, social, and economic contexts. Assuming not only that the perceptuo-motor and memory systems are built to represent the statistical structure of imbalanced environments (Anderson, 1990; Shepard, 1994/2001) but also that the cognitive system has been similarly constructed, we have proposed QuickEst, a fast and frugal

How Estimation Can Benefit From an Imbalanced World

heuristic for making estimations. Its architecture exploits the world's frequent predictable imbalance. In the study of mental tools (including heuristics) as well as mental structures (including perception and memory) we begin to discern that the mind looks very much matched to key structures of the world.

Notes:

(1.) Of course, this line is by definition downward sloping (because the *rank* variable represents a transformation of the *fortune* variable that entails a negative correlation between the two variables). The fact that one observes a straight line, however, is not trivial because there is no tautology causing the data to automatically follow a straight line. As Newman (2005) pointed out, few real-world distributions follow a power law over their entire range. This is particularly true for smaller values of the variable being measured or for very large values. In the distribution of city sizes, for instance, the political capitals, say Paris or London, are much larger than the line drawn through the respective distribution of cities would lead one to expect—they are “essentially different creatures from the rest of the urban sample” (Krugman, 1996). In Figure 15-1b, the 30 richest billionaires' wealth deviates from the fitted straight line: Their wealth is less large than theoretically expected.

(2.) Zipf's law and the Pareto distribution differ in several respects (see Newman, 2005). Pareto was interested in the distribution of income and asked how many people have an income greater than x . The Pareto law is given in terms of the cumulative distribution function; that is, the number of events larger than x is an inverse power of x : $P(X > x) \propto x^{-k}$. In contrast, Zipf's law usually refers to the size y of an occurrence of an event (e.g., the size of a city or the frequency of use of a word). Another difference is the way the distributions were plotted: Whereas Zipf made his plots with rank on the horizontal axis and size on the vertical axis, Pareto did it the other way round.

(3.) When the heuristic is initially set up, only as many cues (of all those available) will be used in the cue order as are necessary to estimate the criterion of four-fifths of the objects

How Estimation Can Benefit From an Imbalanced World

in the training set. The remaining one-fifth of the objects will be put in the catchall category.

(4.) By building in spontaneous numbers, the heuristic models the observation that when asked for quantitative estimates (e.g., the number of windmills in Germany), people provide relatively coarse-grained estimates (e.g., 30,000, i.e., 3×10^4 , rather than 27,634). Albers (2001) defined spontaneous numbers as numbers of the form $a \times 10^i$, where $a \in \{1, 1.5, 2, 3, 5, 7\}$ and i is a natural number.

(5.) In fact, when the training set (100%) equals the generalization set, the estimation tree achieves the optimal performance. Specifically, the optimal solution is to memorize all cue profiles and collapse cities with the same profile into the same size category. In statistics, this optimal solution is known as true regression. Under the circumstances of complete knowledge, the estimation tree is tantamount to true regression.

(6.) There are different definitions of scarcity of information. In the present analysis, we define scarcity in terms of the number of objects on which a strategy is trained compared to the total number of objects in an environment (on which the strategy can be tested). Martignon and Hoffrage (1999, 2002) defined information scarcity in terms of the ratio of the number of binary cues to the number of objects in an environment.

(7.) The number of objects per cue is a poor predictor of QuickEst's performance in relation to that of the estimation tree (regardless of whether the fish fertility environment is included in the analysis).

(8.) We are grateful to Bettina von Helversen and Jörg Rieskamp for their valuable input on the following sections.

(9.) Inferences from givens (i.e., using displayed information) are an unsuitable test-bed for memory-based heuristics like QuickEst. Inferences from givens do not invoke the costs associated with search in memory—including cognitive effort, time, and opportunity costs—which are likely to be key triggers for the use of QuickEst and other heuristics (e.g.,

How Estimation Can Benefit From an Imbalanced World

Bröder & Schiffer, 2003b; see also chapter 9). Hausmann and colleagues (2007; Experiment 2) and von Helversen and Rieskamp (2008) tested QuickEst in the unsuitable context of inferences from givens.



Access brought to you by: Max-Planck Society