Fast and frugal forecasting

Daniel G. Goldstein [a,*], Gerd Gigerenzer [b]

[a] *London Business School, UK*
[b] *Max Planck Institute for Human Development, Germany*

Abstract

Simple statistical forecasting rules, which are usually simplifications of classical models, have been shown to make better predictions than more complex rules, especially when the future values of a criterion are highly uncertain. In this article, we provide evidence that some of the fast and frugal heuristics that people use intuitively are able to make forecasts that are as good as or better than those of knowledge-intensive procedures. We draw from research on the adaptive toolbox and ecological rationality to demonstrate the power of using intuitive heuristics for forecasting in various domains including sport, business, and crime.

© 2009 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

Keywords: Heuristics; Fast and frugal heuristics; Unit weighting; Robustness; Overfitting; Cross-validation

1. Background

1.1. Distrust of simplicity

Science is based on intuitions and would not advance without them. However, some intuitions can block progress, such as the belief that complex problems need complex solutions. In 1979, Makridakis and Hibon tested 22 forecasting models on 111 time series from business and economics and reported that a very simple model (one that basically only weights the most recent observations) made better predictions than more complex models. The complex models were better at fitting the data, but, as an unfortunate consequence, suffered from overfitting. The comments on this finding, published in the *Journal of the Royal Statistical Society* (Makridakis & Hibon, 1979), ranged from praise for conducting such a demanding study, to outright disbelief in the results, to distrust in the authors' competence in performing time-series analyses. In reaction, Makridakis and others conducted further competitions in which outside experts were invited to carry out the forecasts, and reconfirmed the finding that complex methods do not always provide better forecasts than simpler ones (Makridakis et al., 1982, 1993; Makridakis & Hibon, 2000). Did these surprising results revolutionize forecasting methods and generate systematic analyses of when simplicity pays? Was the wisdom of simple averaging as a forecasting

* Corresponding address: London Business School, Regent's Park, Sussex Place, London, NW1 4SA, UK. Tel.: ++44 20 7000 8611.

E-mail address: dgoldstein@london.edu (D.G. Goldstein).

technique investigated? No. At that time, it led to virtually no theoretical work (Fildes & Makridakis, 1995):

"...the evidence is straightforward: Those interested in applying forecasting regard the empirical studies as directly relevant to both their research and application ...those interested in developing statistical models ...pay little attention or ignore such studies" (p. 300).

According to Fildes and Nikolopoulos (2006), who interviewed Makridakis, the situation has not changed much today.

The case of Makridakis is not an exception. In the 1970s, several researchers (Dawes, 1979; Dawes & Corrigan, 1974; Schmidt, 1971) showed that a linear model with equal weights (or random weights of the correct sign) could predict various criteria, such as students' grade point averages, about as well as multiple regression could. Einhorn and Hogarth (1975) asked in which environments equal weights are as good as or better than regression weights, and concluded that typical preconditions include moderate to low linear predictability (R^2 of 0.5 or smaller) and correlated predictors. However, this result has had virtually no influence on the routine use of multiple regression in the social sciences, or on the belief that beta weights are indispensable for good prediction. In an analysis of econometric textbooks, Hogarth (in press) found not a single citation or mention of the predictive power of equal weights. Similarly, in judgment and decision research, the fact that a person ignores information is taken a priori (typically without an empirical check) as indicative of a reasoning error (e.g., Conlisk, 1996). Since cognitive heuristics nearly always ignore information, the phrase "heuristics and biases" has become something close to tautological. In this view, relying on all available information and using sophisticated algorithms to combine information appears to be a sign of rationality, whereas unit-weighting and ignoring information appears irrational.

In this article, we propose that, like simplified statistical models, some of the fast and frugal heuristics people use intuitively can predict the future about as well as sophisticated forecasting models can. By way of introduction, we use a temperature prediction problem to illustrate this point.
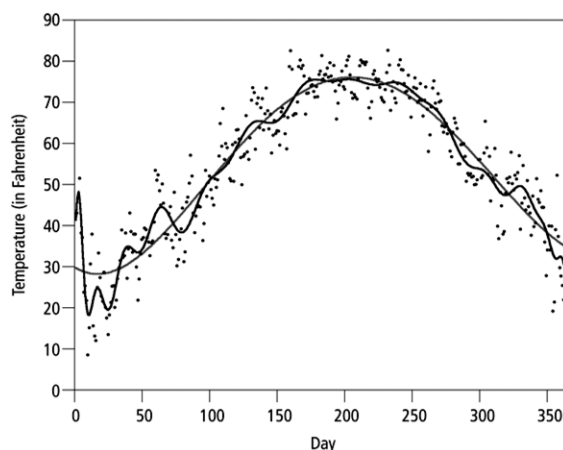

Fig. 1. Average temperature in New York for each day in 2004. Shown is the best-fitting 4th-degree polynomial (smooth curve) and the best-fitting 25th-degree polynomial (jagged curve).

1.2. Best fit does not mean best prediction

Consider the task of trying to predict next year's daily temperatures based on this year's numbers. Fig. 1 shows the average temperature in New York for each day in 2004. To make a family of forecasts for 2005, we fit polynomials of increasing degrees to the 2004 temperatures. A first-degree polynomial has the form $y = w_1 x + c$, a second-degree polynomial has the form $y = w_1 x + w_2 x^2 + c$, and so on, where y is the estimated temperature of the model and w is a vector of free parameters to be fitted to the data. A first-degree polynomial is a straight line, a second-degree polynomial is a parabola. Since neither of these can capture periodicity, both can be excluded as reasonable models of temperature over the year. Fig. 1 shows the fit of both a fourth-degree and a 25th-degree polynomial to the data. It is easy to see that the 4th degree polynomial fits worse than the 25th degree one, and, as it turns out, the average errors (absolute difference) are about 5.1 and 4.3 °F (or 2.8 °C and 2.4 °C), respectively. Fig. 2 shows the fit of all polynomials, indicating, as one might guess, that the more complex the polynomial, the better the fit (note that lower is better). In many articles in the cognitive sciences, sociology, and economics, the best fit is taken as evidence for the best model, and in such a view, complex models are preferred (e.g., see Roberts & Pashler, 2000).
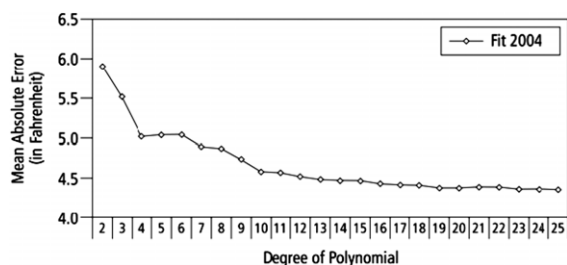
Fig. 2. How well can polynomials of different degrees fit the average temperature in New York, 2004? Shown is the fit of all polynomials of degree 2 to 25, including that of the two polynomials in Fig. 1. (The fit of the 1st-degree polynomial is not shown, since it is out of the scale.)

But does the model with the best fit also give the best prediction of next year's temperature? To answer this question, we took each polynomial (with the optimal weights fitted) and predicted the next year's temperatures. Call this the forecasting curve. Before you turn the pages, please test your own intuition about what shape this curve will have. Will it be above or below the fitting curve in Fig. 2? Will it decrease from left to right, like the fitting curve? Will it be horizontal? Will it increase from left to right? Will it cross the fitting line? We have asked these questions in many seminars of a diverse range of audiences from cognitive scientists to economists. Almost everyone surmises that the forecasting performance curve will be above the fitting curve, because foresight is more difficult than hindsight. But few guess what Fig. 3 shows: the forecasting performance first gets better (i.e., error decreases) over the leftmost three points, but then gets progressively worse as the error gradually increases to the right. The best forecasting performance was obtained with the relatively simple 4th-degree polynomial, whereas a further increase in free parameters led to increasingly worse predictive accuracy. When model A has a better fit than model B, but B forecasts better than A, we say that A overfits the data relative to B. On the other hand, the 2nd- and 3rd-degree polynomials have both a lower fit and a lower forecasting accuracy than the 4th-degree polynomial. This is called underfitting, extracting too little information from the historical data. A model is called robust relative to the degree that it retains accuracy when its fitting and forecasting performance are compared.

By experimenting with different years or cities, anyone unfamiliar with robustness can learn first-hand that when the criterion cannot be perfectly predicted, there is a point where more complexity hurts. Simplicity helps, or, as Einstein (1934) said, "the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience" (p. 165). The popular paraphrase of Einstein is "make things as simple as possible, but not simpler", reflecting, perhaps, that the populace has taken the message to heart.

1.3. The adaptive toolbox

The study of the adaptive toolbox concerns the heuristics people use for judgment and decision making, and the analysis of their building blocks. Through recombination, these building blocks allow for the creation of heuristics that are adapted to new environments (Gigerenzer & Selten, 2001; Gigerenzer, Todd & The ABC Research Group, 1999). Cognitive heuristics are strategies that humans and other animals use. We call them fast because they involve relatively little estimation and frugal because they ignore information. A heuristic is not either good or bad per se. Its performance is dictated by features of the *information environment*, such as low predictability, or high cue redundancy. The study of ecological rationality is the study of how information environments cause heuristics to succeed or fail. Table 1 defines the subset of heuristics that are used in this article for forecasting purposes, including examples of conditions under which they perform admirably relative to more complex strategies (Gigerenzer, 2008; Goldstein et al., 2001).

Because heuristics are fast and frugal, they have been perceived in some areas of psychology as second-best strategies, the necessary outcome of our cognitive limitations; but, in this tradition, there are few computational models of heuristics. By formulating and testing computational models of heuristics, however, it has been shown that (i) people's behavior is often better explained by models of heuristics than by complex information processing steps, such as the weighting and summing of information (e.g., Bergert & Nosofsky, 2007) or exemplar models of categorization (Nosofsky & Bergert, 2007), and (ii) provided that people have
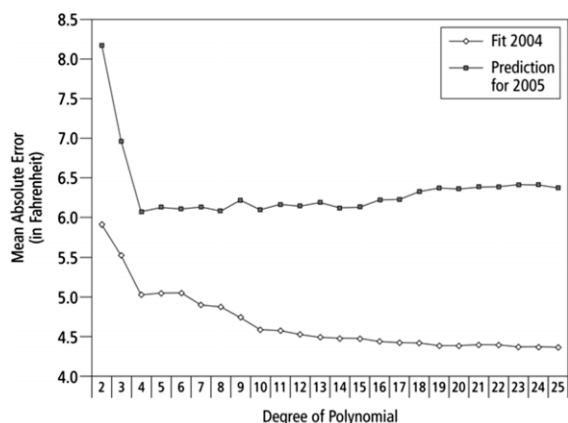
Fig. 3. How well can the best-fitting polynomials predict the average temperature in New York, 2005? The upper curve shows that the error of the 4th-degree polynomial is lowest, whereas the polynomials of higher degree overfit the 2004 data, and those with lower degree underfit the data.

sufficient accurate feedback, heuristics are used in an adaptive way. That is, heuristics are used in environments where they are ecologically rational (e.g., Dieckmann & Rieskamp, 2007; Rieskamp & Otto, 2006). Interestingly, some heuristics seem to cross species boundaries; for instance, one-good-reason decisions, such as those modeled by take-the-best, have been observed in both animals and humans (see Hutchinson & Gigerenzer, 2005).

In this article, we provide examples of some cognitive heuristics that can do reasonably well at predicting the future. The underlying logic is the same as in Fig. 3: in an uncertain world, an organism has to find ways to detect the important cues and ignore the rest. In statistical terms, when faced with out-of-sample or out-of-population prediction, a forecasting method has to bet on robustness instead of attempting to secure an optimal fit to the past, particularly if samples are small, cues are abundant, predictability is moderate or low, and there is a chance of overfitting. We turn now to examine simple heuristics in the domains of sport, business, and crime.

2. Domain: Sport

2.1. Predicting Wimbledon

Every year, millions of spectators watch the tennis matches at Wimbledon, one of the four annual "grand slam" tennis events, and the only one still played on natural grass. In the Gentlemen's Singles Championship, 128 players compete in 127 matches. (The number of matches is 127 because one player is eliminated from the tournament with every match, except the champion who never loses a single game.) The forecasting problem here is to predict the outcomes of all matches before the tournament begins.

2.1.1. Competitors

Three of the four competing forecasting strategies have extensive information about all contestants: (1) the ATP (Association of Tennis Professionals) Champions Race, the official worldwide ranking of tennis players for the calendar year; (2) the ATP Entry Ranking, the official ranking for the last fifty-two weeks; and (3) the seeding of players, which represents the expert ranking of the Wimbledon officials. The experts know the ATP rankings, but they typically deviate from them, taking into account their specific expert knowledge about the players and the specifics of the Wimbledon tennis court, such as the players' success in grass tournaments. World rankings and expert seeding have been shown to be good predictors of sport outcomes (Boulier & Stekler, 1999, 2003). For each strategy, the same decision rule was used: the player with the higher rank will win the game. The final forecasting strategy is a heuristic rule that is a simple extension of the recognition heuristic (Table 1):

Collective recognition heuristic: Ask a sample of semi-informed people to indicate whether they have heard of each player or not. Rank players according to collective recognition, and predict, for each match, that the player with the higher rank will win.

The term "semi-informed" refers to people who recognize the names of *some* of the players, but not all. Why is being only somewhat informed important? The recognition heuristic can only be applied if the names of only some of the players are known; that is, experts who have heard of all players cannot use it, and it can be used most often if a person has heard of about half of the players (Goldstein & Gigerenzer, 2002).

2.1.2. Results

There are two tests of the forecasting accuracy of collective recognition. In the first test conducted by

Table 1
Cognitive heuristics used in this article for forecasting.

Heuristic	Definition[a]	Conditions favoring (relative) performance:	Bold predictions and results
Recognition heuristic (Goldstein & Gigerenzer, 2002)	If one of two alternatives is recognized, infer that it has the higher value on the criterion	Recognition validity >0.5	Contradicting information about recognized object is ignored, less-is-more effect if $\alpha > \beta$, forgetting is beneficial
Fluency heuristic (Schooler & Hertwig, 2005)	If one alternative is recognized faster than another, infer that it has the higher value on the criterion	Fluency validity >0.5	Less-is-more effect, forgetting is beneficial
Take-the-best (Gigerenzer & Goldstein, 1996)	To infer which of two alternatives has the higher value: (a) search through cues in order of validity, (b) stop search as soon as a cue discriminates, (c) choose the alternative this cue favors	Cue validities vary highly, moderate to high redundancy, scarce information (Hogarth & Karelaia, 2005, 2006; Martignon & Hoffrage, 1999, 2002)	Made predictions equally or more accurate than regression (Czerlinski, Goldstein, & Gigerenzer, 1999), neural networks, exemplar models, and CARTs (Brighton, 2006)
Tallying (equal-weight linear model; Dawes, 1979)	To estimate a criterion, do not estimate weights but simply count the number of favoring cues	Cue validities vary little, low redundancy (Hogarth & Karelaia, 2005, 2006)	Can predict as accurately as or more accurately than multiple regression
Hiatus heuristic (Wübben & Wangenheim, 2008)	Assume that customers who have not purchased in a fixed period of time are inactive	Not investigated	Performed as well as the Pareto/NBD model
Persistence of best customers (Wübben & Wangenheim, 2008)	Assume that the best X% of customers in the past will be the best X% of customers in the future	Spare	Performed as well as the Pareto/NBD and BG/NBD models
$1/N$ heuristic (DeMiguel, Garlappi, & Uppal, 2009)	Allocate resources equally to each of N alternatives	High unpredictability, small learning sample, large N	None of 14 optimal asset allocation models consistently outperformed it
Center-of-the-circle heuristic (Snook, Zito, Bennell, & Taylor, 2005)	Predict that the criminal lives at the midpoint of the two farthest apart crimes	Number of crimes in sequence less than 9	Made better forecasts of location than 10 complex models

[a] For formal definitions, see references.

Serwe and Frings (2006), the collective recognition of a group of German amateur players who had heard of about half of the contestants was obtained before the beginning of the 2003 tournament, and the same was done for a second group of laypeople who only had heard, on average, of 14 of the players. Fig. 4 shows that the two ATP rankings predicted the winners correctly in 66% and 68% of the matches, respectively. The experts did slightly better. Their seeding predicted the outcomes of 69% of the matches correctly. Yet the collective recognition rule predicted 66% for laypeople with very low tennis player name recognition, which was as good as the ATP Entry Ranking did, while the collective recognition of the semi-informed respondents reached 72% correct.

Note that the good performance of recognition was obtained despite the participants having a distinct handicap: they recognized more German players, who fared poorly in this tournament.

This study found that laypeople and amateurs made forecasts consistent with the recognition heuristic in 90% of the cases in their individual predictions. To test whether the fact that recognition beat the three benchmarks was a lucky event, another group
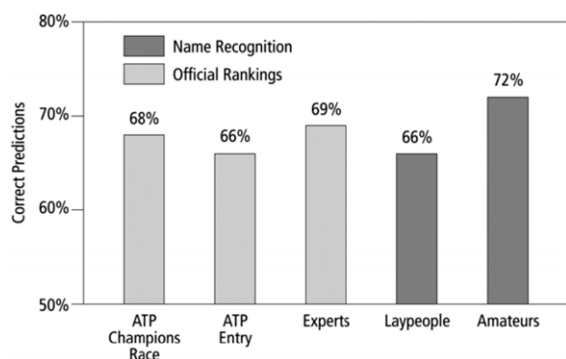
Fig. 4. Predictions of the outcomes of the 2003 Gentlemen's Singles tennis matches. Benchmarks are the ATP Champions Race, the ATP Entry Ranking, and the seeding of the Wimbledon experts. The collective recognition of laypeople who had heard of only a few players and amateurs who had recognized only half of them by name predicted the outcomes as well as or better than the official benchmarks (data from Serwe & Frings, 2006).

of researchers used essentially the same method to predict the results of the Gentlemen's Singles Championship in 2005 (Scheibehenne & Bröder, 2007). They reported that in 2005, the ATP Entry Ranking, the ATP Champions Race and the seedings of the Wimbledon experts led to 69%, 70%, and 70% of correct forecasts. The collective recognition of lay people, amateur players, and all participants combined led to 67%, 68%, and 70% correct, respectively. These authors also quantitatively analyzed the question of why mere recognition is such a powerful predictor, by analyzing the structure of the environment that generates recognition without being an expert in tennis. The Kruskal's gamma correlation between each player's success (the number of matches won in the 2005 tournament) and how often this player was mentioned in a major newspapers (where the participants lived) was 0.33, the correlation between the frequency of mentions in the newspapers and collective recognition was 0.58, and the correlation between recognition and success was 0.40. This analysis shows in quantitative terms that there is information in a lack of recognition, and that collective wisdom can emerge from aggregated individual semi-ignorance. It also provides a way to gauge when collective recognition will be successful and when it will not.

These two studies indicate that if there is a correlation between the criterion and name recognition, then a beneficial degree of ignorance can forecast the winners in the Wimbledon tennis matches as well as a weighted record of the players' performance and the seeding of the Wimbledon experts.

2.2. Forecasting soccer

In the tradition that expertise and sophisticated knowledge are needed to predict the outcomes of uncertain events, the study of sports forecasting has primarily focused on the forecasts of experts (e.g., Forrest & Simmons, 2000). Anderson, Edman, and Ekman (2005) posed the following question: How much better can soccer experts (sport journalists, soccer coaches and soccer fans) predict the outcome of the first round of the World Cup 2002 than non-experts? In the first round, there are 32 teams in 8 groups, and two in each group move on to the second round. Before the start of the Cup, a total of 251 experts, knowledgeable Swedish students, naïve Swedish students, and American students were asked to forecast which two in each group will move into the next round, and to rate their confidence in their forecast and their knowledge about each team. The median number of correct forecasts was 9 for the experts, compared to 9.5, 10, and 10, for the knowledgeable Swedish, naïve Swedish, and American students. When in a second condition, students were given information about the teams, it had no impact on predictive accuracy. Note that both experts and laypeople consistently predicted slightly better than chance (8 correct). How could laypeople, including American students with extremely little knowledge about soccer (much less European soccer), predict as well as and even somewhat better than the experts? The authors did not directly test models of heuristics, but report that, in the majority of cases, participants picked the two teams from each group that they knew best, and relate this to the recognition heuristic.

A similar observation was made by Ayton and Önkal (1997), who reported that Turkish business students forecasted the outcomes of 32 English FA Cup matches almost as well as knowledgeable British soccer fans (62.5% vs. 65.6%), despite their knowing very little about British teams. The Turkish students achieved this performance by consistently predicting (that is, 93% of the time) that the team which was

more familiar to them would win the match. Once again, this may be related to the recognition heuristic, or the fluency heuristic (Table 1). As with the previous study, no direct test of the forecasting power of either heuristic was conducted. Pachur and Biele (2007) conducted a direct test of mere recognition to predict the winners of the 24 first-round matches of the 2004 European Soccer Championship. In contrast to the previous results, forecasts based on laypeople's recognition were less accurate than the average expert's forecasts, whereas the FIFA rank (world rank of soccer teams) and the performance in the qualifying round outperformed the experts. It is worth noting that this result may differ from the previous one because Greece won, despite being an extreme long-shot that few laypeople recognized. Gröschner and Raab (2006) asked 208 experts and laypeople to predict the 2002 soccer world champion. Laypeople did substantially better, predicting the winner twice as often as experts. Many laypeople relied on a simple heuristic, bet that the team who had won most championships beforehand will win it again (Brazil), and this time they were right.

There is a need to test the relative performance of heuristics, experts, and complex forecasting methods more systematically over the years rather than in a few arbitrary championships (Bennis & Pachur, 2006). Overall, though, the results indicate that collective and individual recognition, or fluency, can predict the winners of Wimbledon and soccer championships about as well as expertise based on substantial knowledge.

3. Domain: Business

3.1. Forecasting future purchase activity

In an age in which companies maintain databases of their customers' historical purchase data, a key problem becomes predicting which customers are likely to purchase again in a given time frame, and which are inactive. Armed with this information, managers can make decisions about where to spend their limited marketing budgets. How well can a simple managerial rule do compared to a knowledge-intensive probability model when it comes to classifying customers as active or inactive?

3.1.1. Competitors

Wübben and Wangenheim (2008) interviewed managers in different industries and characterized their classification of customers with a simple rule, which they formalized as the "hiatus heuristic":

Hiatus heuristic: If a customer has not purchased within a certain number of months in the past (the "hiatus"), the customer is classified as inactive, and otherwise active.

The authors staged a prediction competition. Pitted against the hiatus heuristic was the Pareto/NBD model from the marketing literature, which assumes that purchases follow a Poisson process with a purchase rate parameter λ, that customer lifetimes follow an exponential distribution with a dropout rate parameter μ, and that, across customers, purchase and dropout rates are distributed according to a gamma distribution. In all, four parameters need to be estimated: two (r and α) for the gamma distribution of the purchase rates over individuals, and two (s and β) for the gamma distribution of dropout rates. The Pareto/NBD model also has a complex likelihood function associated with it, requiring multiple draws from a Gaussian hypergeometric function, which is both computationally demanding and unfamiliar to the majority of practitioners (Fader, Hardie, & Lee, 2005).

Customer data from the airline, apparel and online music industries were used to test the models. The apparel data, for example, contained initial and repeat purchase information for 2330 customers of a retailer over a period of 80 weeks. The Pareto/NBD estimated its four parameters on the first 40 weeks of data and tested on the latter 40. For the hiatus heuristic, no parameters were estimated from the data. Instead, interviews with marketing managers were conducted to obtain hiatus values for the apparel and airline industries, and a reasonable guess was made for the online music business.

3.1.2. Results

For the apparel data, the hiatus heuristic correctly classified 83% of the customers, while the Pareto/NBD model achieved 75% correct. For the airline data, the score was 77% vs. 74%, and in the online music data, the two methods tied at 77%. In a second competition, which attempted to find the optimal thresholds for both the heuristic and Pareto/NBD

Table 2
Asset allocations models tested by DeMiguel et al. (2009).

Number	Model	Category
0	$1/N$ heuristic	Simple heuristic
1	Sample-based mean-variance	Classical approach, ignores estimation error
2	Bayesian diffuse prior	Bayesian approach to estimation error
3	Bayes–Stein	Bayesian approach to estimation error
4	Bayesian data-and-model	Bayesian approach to estimation error
5	Minimum variance	Moment restriction
6	Value-weighted market portfolio	Moment restriction
7	MacKinlay and Pastor's (2000) missing factor model	Moment restriction
8	Sample-based mean variance with shortsale constraints	Portfolio constraints
9	Bayes–Stein with shortsale constraints	Portfolio constraints
10	Minimum variance with shortsale constraints	Portfolio constraints
11	Minimum variance with generalized constraints	Portfolio constraints
12	Kan and Zhou's (2007) "three fund" model	Optimal-combinations of portfolios
13	Mixture of minimum-variance and $1/N$	Optimal-combinations of portfolios
14	Garlappi, Uppal, & Wang (2007) multi prior model	Optimal-combinations of portfolios

model, the surprising result emerged that the heuristic made slightly more accurate forecasts in all three domains.

Why does the hiatus heuristic succeed? Without needing to estimate parameters, the heuristic avoids wildly inaccurate results that may arise when parameters are mis-estimated on relatively small samples. In effect, its training has happened in the past, and it is likely to have improved as it was passed from manager to manager. If the wrong hiatus length is chosen, the heuristic could fail. However, it turns out in these analyses that the intuitions of actual managers about the hiatus length were surprisingly close to optimal. For example, for the airline and apparel industries, the manager's intuitions for the length of the hiatus were 3 quarters and 39 weeks, respectively. The optimal figures turned out to be 4 quarters and 40 weeks. For these two industries, choosing the optimal hiatus over that from the managers' intuitions would yield an improvement of less than 1% in terms of correctly classified customers.

3.2. Forecasting future best customers

Marketing managers want to forecast who their future high-value customers will be. They do so for many reasons, in particular to give them "best customer" treatment that may prevent them from switching to a competitor (Malthouse & Blattberg, 2005). Since past best customers are not necessarily future best customers, this would appear to be a complex task. As often happens, complex probability models have been applied to the challenge.

3.2.1. Competitors

Wübben and Wangenheim (2008) have tested the Pareto/NBD model (described in the previous section), in addition to a somewhat simpler variant called the BG/NBD model (Fader et al., 2005), on the task of predicting from historical data which customers will emerge as the best customers in the future. Both classes of model estimate four parameters on training data. In contrast, the managerial heuristic examined estimates no parameters. It simply predicts that the top X% of customers in the past will continue to be the top X% of best customers in the future.

For the same airline, apparel, and online music industry datasets, customers were categorized as either in or out of the top 10% or 20% by number of transactions. The key statistic of managerial interest in this domain is the percentage of predicted best customers who actually turned out to be best customers. On this criterion, out of 12 tests (2 competitor models, 3 industries, and 2 levels of best customer being top 10% or 20%), the heuristic beat the stochastic models all but 3 times. Further analyses showed that varying the length of the holdout set had little impact on the results.
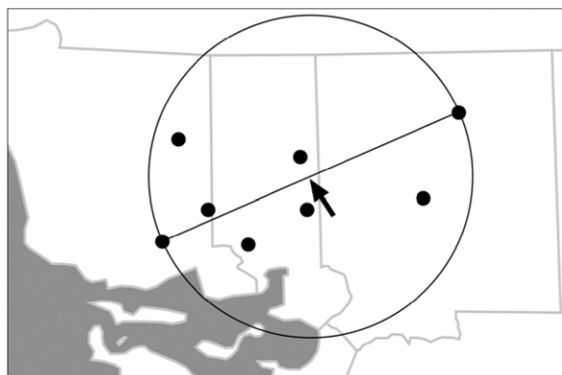
Fig. 5. The center-of-the-circle heuristic predicts that the serial criminal lives near the center of the circle encompassing all of crimes in the series. This point, marked with the arrow, is simply the midpoint of the line connecting the two farthest apart crime scenes.

3.3. Choosing the best portfolio

Professional investors, as well as laypeople investing for their own retirement, make asset allocation decisions; that is, they must choose which investments to hold and how much wealth to allocate to each. The performance of portfolios is often judged with a ratio of reward to volatility, such as the Sharpe ratio, or with other metrics such as the certainty-equivalent return (CEQ). DeMiguel et al. (2009) asked which of 14 methods of portfolio construction would lead to the highest Sharpe ratio, CEQ, and lowest turnover, when tested on 6 historical data sets.

3.3.1. Competitors

Table 2 lists the asset allocation models considered. The spirit of the simplest heuristic in this competition can be found in the fourth century Babylonian Talmud, which states "One should always divide his wealth into three parts: a third in land, a third in merchandise, and a third ready to hand" (Tractate Baba Mezi'a, folio 42a). This simple rule is called the *1/N heuristic*, and in recent times it has been described as a naïve strategy used by lay investors (Benartzi & Thaler, 2001). At each rebalancing, this heuristic simply puts $1/N$ of wealth into each of the N available investment options. For the dataset of the S&P 500, the N options considered were the ten industry sectors (such as financial, healthcare, energy, etc.), in addition to the US equity market portfolio, making an N of 11. The $1/N$ heuristic does not estimate any parameters based on training data.

The competitors, numbered 1 to 14 in Table 2, include both Bayesian and non-Bayesian optimizing models. For such models to work, considerable estimation from a training set is necessary. For instance, for the mean-variance model, a vector of expected excess returns and a variance-covariance matrix need to be estimated. Estimation helps models by allowing them to adapt to the peculiarities of different datasets; however, the downside of estimation is the possibility of estimation error. In fact, many of the models listed have come into existence to deal with estimation error.

3.3.2. Results

Across seven empirical datasets, none of the 14 optimizing models is consistently better than $1/N$ heuristic in terms of the Sharpe ratio, certainty-equivalent return, or turnover. For instance, when looking at Sharpe ratios, in only 6 out of 72 cases did one of the optimizing models outperform the $1/N$ heuristic by a statistically significant margin. For the certainty-equivalent measure, this occurred in only 2 of the 72 tests. When it came to turnover, only one strategy – holding the market and not trading – was better than the $1/N$ heuristic.

What are the conditions under which the heuristic does well relative to the optimizing models? DeMiguel, Garlappi and Uppal list two key factors. The first is a large N. Not only is diversification achieved when money is spread among many investments, but optimizing models suffer when estimating large sets of parameters. The second factor is a short time horizon. The amount of data needed for the accurate estimation of weights can be very long. With typical parameters, and 25 assets to invest in, about 250 years of training data are needed for the sample-based mean-variance method to beat $1/N$, and 500 years are needed with an N of 50. In several hundred years, the time may come to switch from simple to "optimal" asset allocation techniques — assuming that the same stock markets are still around.

4. Domain: Crime

When a number of crimes, for instance burglaries, can be linked to the same offender, police often plot their locations on a map. The art of finding the location of the criminal's home based on the crime sites is a key objective in what is known as *geographical profiling*.

Table 3
Accuracy of 11 geographic profiling strategies. Descriptions, data, and strategies adapted from Snook et al. (2005), which should be consulted for details.

Strategy	Description	Mean error (kilometers from offender's home)
Center of the circle heuristic	Predict the midpoint of the line connecting the two farthest apart crimes	8.01
Centroid	Point whose coordinates are the mean of the x and y coordinates of the crime sites	8.07
Harmonic mean	Point whose coordinates are the inverse mean of the inverse coordinates	8.08
Geometric mean	The anti-log of the mean of the logarithms of the coordinates	8.08
Center of minimum distance	The point in a grid where the sum of the distance between that point and all crime locations is smallest	8.33
Median	The middle value of the distribution of coordinates	8.46
Linear	The probability of an offender living at a particular location decreases in a linear fashion with increasing distance away from a crime site	8.47
Normal	Assumes the likelihood of the offender's home location peaks at some optimal distance from the crime sites then declines as a normal distribution	8.81
Lognormal	Same as above but with skewed lognormal distributions	8.94
Negative exponential	Assumes likelihood of home location is highest at a crime site, decreasing exponentially with distance	9.03
Truncated negative exponential	Combines the linear strategy and the negative exponential strategy	9.06

4.1. Competitors

Snook et al. (2005) ran a competition between 11 techniques for locating offender residences. All techniques used as input the $x - y$ coordinates of crimes (by one serial criminal) on a map and made predictions of the position of the criminal's home. There are many ways to turn these sets of coordinates into a point prediction. As with the other examples, one stands out as exceptionally simple, so much so that it can be carried out with a pencil and ruler:

Center of the circle heuristic: Predict that the offender lives at the mid-point of the line connecting the two farthest apart crime locations.

Fig. 5 shows the heuristic applied to a set of crime locations. Ten other methods for profiling were tested, including other "spatial distribution strategies" such as finding the centroid, harmonic mean, geometric mean, or point of minimum distance. Also investigated were computationally intensive "probability distance strategies" that involve fitting probability distributions such as the negative exponential, normal and lognormal. To carry out these models, a computer program superimposes a grid (of arbitrary granularity) over a map containing the n crime locations. From the center $\underline{C_i}$ of each of the $\underline{i}$ cells in the grid, the distance $\underline{dist_{i,j}}$ to the $\underline{j}$th crime location is computed.

A probability density function f is used to give a likelihood to each cell. For example, if the lognormal model is being applied, $l(C_i) = \sum_{j=1}^{n} f(dist_{i,j})$ and $f(x) = \dfrac{a e^{\frac{(2 \log x - \mu)^2}{2\sigma}}}{x^2 \sqrt{2\pi\sigma}}$, where the parameters a, μ and σ are input by the user. The home location of the offender would be predicted to be at the center of the cell that maximizes the function l.

To conduct the competition, the various methods were applied to the crime locations of 16 UK residential burglars who had committed at least 10 crimes. Based on the first 5 to 10 crime locations, the Euclidian distance between the forecasts and the burglar's home was computed. The simplest spatial distribution strategy (the centroid) takes 22 computational steps to make a forecast from 5 crimes, while the simplest probability distance method takes 85,625. Interestingly, all of the strategies were run on a computer except for the center-of-the-circle heuristic, which was applied manually. What is simple for a human may be complex for a computer, and vice versa.

4.2. Results

As is shown in Table 3, across all series lengths, the center-of-the-circle heuristic made the most accurate forecasts of where the criminal lived. The authors found the best strategy when the training data included

from 5 to 10 crimes. In four out of these six conditions, the center-of-the circle heuristic came in first out of 11 methods. On average, it beat even the best-performing complex strategy by 1.25 km. Only when there was a large sample size of burglaries known to be committed by the same criminal, such as a "training set" with 9 and 10 crimes, did the simple circle heuristic fall in the rankings, but at this point the difference between the best and worst methods was only 0.25 km. In a learning study (Snook, Taylor, & Bennell, 2004), laypeople who were trained with heuristic methods were as accurate at predicting home locations as a computerized geographic profiling system. As with sports and business, forecasting where criminals live can be efficiently handled with fast and frugal heuristics.

5. Conclusion

It seems that humans and other animals have always relied on fast and frugal heuristics. To measure the area of a potential nest cavity, an ant has no yardstick but a rule of thumb: run around for a fixed period, leave a pheromone trail, go away, come back, move around on a different irregular path, and estimate the size of the cavity by the frequency of running into the old trail. This heuristic gives remarkably precise estimates (see Hutchinson & Gigerenzer, 2005). Yet, for some reason, in human psychology, heuristics are often interpreted as cognitive flaws. Cognitive scientists tend to recreate humans in the image of God, just as the Old Testament says, assuming some form of omniscience (knowing all probabilities) and omnipotence (being able to compute all functions in a split second). We know now that the false intuition "complex problems require complex solutions" can actually block progress in statistical models of forecasting (e.g., Makridakis & Hibon, 1979). As we have shown in this article, the same result holds for the heuristics humans use. Less can be more.

To forecast a criterion that is not perfectly predictable, which is almost always the case, a method has to be robust. Fast and frugal heuristics achieve robustness through their simplicity and sparing use of information. But not every heuristic can predict the future equally well, and the problem for contemporary researchers has become understanding which heuristic is best suited to which problem. The answers to this question are found in the study of the ecological rationality of heuristics (Table 1; Goldstein et al., 2001, Hogarth & Karelaia, 2006). Consider how certain phenomena we have discussed resemble the wisdom of crowds (Surowiecki, 2004), and others the wisdom of individuals, laypeople, and experts. If we can specify the processes underlying the wisdom of collective intelligence, we can begin to understand when it succeeds and when it fails.

Let us end on a practical note. Simplicity creates not only robustness, but also transparency. The simple heuristics we have shown for investing and geographical profiling, for instance, are transparent and can easily be understood by practitioners, whereas many of the asset allocation models in Table 2 or the probability-distance profiling models in Table 3 remain mystical machinery. The danger is that complex methods become an end in themselves, a ritual to impress others, and at the same time opportunities to learn how to do things better are missed. Learning requires some form of transparency, which forecasters can best achieve when they understand what they are doing.

References

Anderson, P., Edman, J., & Ekman, M. (2005). Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, *21*(3), 565–576.

Ayton, P., & Önkal, D. (1997). Forecasting football fixtures: Confidence and judged proportion correct. City University. London. (Unpublished manuscript).

Benartzi, S., & Thaler, R. (2001). Naive diversification strategies in retirement saving plans. *American Economic Review*, *91*(1), 79–98.

Bennis, W. M., & Pachur, T. (2006). Fast and frugal heuristics in sports. *Psychology of Sports and Exercise*, *7*, 611–629.

Bergert, F. B., & 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*, *33*, 107–129.

Boulier, B. L., & Stekler, H. O. (1999). Are sports seedings good predictors? An evaluation. *International Journal of Forecasting*, *15*, 83–91.

Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *International Journal of Forecasting*, *19*, 257–270.

Brighton, H. (2006). Robust inference with simple cognitive models. In C. Lebiere, & R. Wray (Eds.), *AAAI Spring Symposium: Cognitive science principles meet AI-hard problems* (pp. 17–24). Menlo Park, CA: American Association for Artificial Intelligence.

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

Czerlinski, J., Goldstein, D. G., & Gigerenzer, G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & the ABC Group (Eds.), *Simple heuristics that make us smart*. New York: Oxford University Press.

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

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

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy. *Review of Financial Studies*, *22.5*, 1915–1953.

Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory and Cognition*, *35*, 1801–1813.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, *13*, 171–192.

Einstein, A. (1934). On the method of theoretical physics. *Philosophy of Science*, *1*(2), 163–169.

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). Counting our customers the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, *24*(2), 275–284.

Fildes, R., & Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*, *65*, 289–308.

Fildes, R., & Nikolopoulos, K. (2006). Spyros Makridakis: An interview with the International Journal of Forecasting. *International Journal of Forecasting*, *22*(3), 625–636.

Forrest, D., & Simmons, R. (2000). Forecasting sport: The behaviour and performance of football tipsters. *International Journal of Forecasting*, *16*(3), 317–331.

Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, *20*, 41–81.

Gigerenzer, G. (2008). *Rationality for mortals.*. New York: Oxford University Press.

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

Gigerenzer, G., & Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

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

Goldstein, D. G., Gigerenzer, G., Hogarth, R. M., Kacelnik, A., Kareev, Y., Klein, G., Martignon, L., Payne, J. W., & Schlag, K. (2001). Why and when do simple heuristics work? In G. Gigerenzer, & R. Selten (Eds.), *Bounded Rationality: The Adaptive Toolbox*. Cambridge: MIT Press.

Gröschner, C., & Raab, M. (2006). Vorhersagen im Fußball. Deskriptive und normative aspekte von vorhersagemodellen im sport. *Zeitschrift für Sportpsychologie*, *13*(1), 23–36.

Hogarth, R. M. When simple is hard to accept. In P. M. Todd, G. Gigerenzer, & The ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world*. Oxford: Oxford University Press (in press).

Hogarth, R. M., & Karelaia, N. (2005). Ignoring information in binary choice with continuous variables: When is less "more"? *Journal of Mathematical Psychology*, *49*(2), 115–124.

Hogarth, R. M., & Karelaia, N. (2006). Take-the-best and other simple strategies: Why and when they work 'well' in binary choice. *Theory and Decision*, *61*, 205–249.

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

Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, *42*, 621–656.

MacKinlay, A. C., & Pastor, L. (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of Financial Studies*, *13*, 883–916.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*, 111–153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simmons, L. F. (1993). The M-2 competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*, 5–23.

Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation (with discussion). *Journal of the Royal Statistical Society, Series A*, *142*(2), 79–145.

Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*, 451–476.

Malthouse, E., & Blattberg, R. (2005). Can we predict customer lifetime value? *Journal of Interactive Marketing*, *19*(1), 2–16.

Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.

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

Nosofsky, R. M., & Bergert, F. B. (2007). Limitations of exemplar models of multi-attribute probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 999–1019.

Pachur, T., & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica*, *125*, 99–116.

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

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367.

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, *31*, 699–714.

Serwe, S., & Frings, C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making*, *19*(4), 321–332.

Scheibehenne, B., & Bröder, A. (2007). Can lay people be as accurate as experts in predicting the results of Wimbledon 2005? *International Journal of Forecasting*, *23*, 415–426.

Schooler, L., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, *112*, 610–628.

Snook, B., Zito, M., Bennell, C., & Taylor, P. J. (2005). On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology*, *21*(1), 1–26.

Snook, B., Taylor, P. J., & Bennell, C. (2004). Geographic profiling: The fast, frugal, and accurate way. *Applied Cognitive Psychology*, *18*, 105–121.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York: Doubleday.

Wübben, M., & Wangenheim, F. V. (2008). Instant customer base analysis: Managerial heuristics often get it right. *Journal of Marketing*, *72*, 82–93.