CHAPTER 7

HEURISTICS

DANIEL GOLDSTEIN

I see your face in every flower,
Your eyes in stars above.
It's just the thought of you,
The very thought of you, my love.
('The Very Thought of You')[1]

Introduction

Lovers, it has been remarked, see the faces of those they love in the natural world. Scientists, it has been remarked, fall in love with tools, and see them in the phenomena they investigate (Gigerenzer 1991). For example, one of the most popular statistical tools in the experimental social sciences is analysis of variance, or ANOVA. In the 1970s, a full 70 percent of articles in experimental-psychology journals used this one technique (Edington 1974). In a psychological theory of how the mind attributes a cause to an effect, Kelly (1967) suggested that minds performs causal attribution the same way social scientists do; that is, by means of analysis of variance. The ANOVA theory seems to have resounded well with the scientific audience, racking up more than 900 references in a decade (Kelly and Michaela 1980). Still, today, causal attribution is a leading question is social psychology. Scientific tools transformed into theories again in the years following the introduction of the computer to university departments. In the *cognitive revolution* of the 1970s and 1980s the new field of cognitive science was born based on the idea that the mind is analogous to a computer (Gigerenzer and Goldstein 1996). In the words of a founder of this movement, Herbert Simon (1979: 363), 'The fundamental reactions of the mental chemistry employ elementary information processes that operate upon symbols and symbol structures: copying symbols, storing symbols, retrieving symbols, inputting and outputting symbols, and comparing symbols'. These continue to serve as the basic verbs of cognitive psychology today.

In the field that is the topic of this volume James Coleman proposed the following explanation of why, in the second half of the twentieth century, sociology became decidedly less social:

> One may ask just why there came to be such a radical shift toward a focus on individual behavior in a discipline whose subject matter, after all, is the social system. Part of the answer lies in the invention of techniques. The statistical tools of survey design and analysis began in the 1940s to make possible quantitatively precise statements about samples of independent individuals and the populations (again of independent individuals) they represent, as well as analysis of factors affecting individual behavior. There was no comparable development of tools for analysis of the behavior of interacting systems of individuals or for capturing the interdependencies of individual actions as they combine to produce a system-level outcome. The far greater complexity required of tools for these purposes constituted a serious impediment to their development and continues to do so (though some methods such as those generally labelled 'network analysis' move in that direction). (1986: 1316)

Coleman suggests that tools for analyzing social systems failed to develop because of their complexity. However, some complex tools, such as computers, have both developed at a rapid rate and influenced scientific models. We might take Coleman to mean that it was the complexity of *applying* system-analysis tools that kept them from influencing his field, perhaps due to the reluctance of scientists to adopt them. It has been observed (Gigerenzer 1991; Gigerenzer and Goldstein 1996) that tools are transformed into successful theories only after the tools have achieved widespread adoption. If there ever was a simple, widespread tool in the social sciences, it is the one that we focus on in this chapter: the linear model.

Linear models are everywhere in social science, where they serve as near-universal tools for estimation, modelling, designing experiments, and testing hypotheses. One would be pressed to find a social scientist who has not used ANOVA, multiple regression, the general linear model, or the generalized linear model for statistical analysis. The most frequently used linear models make use of optimization techniques, arriving at coefficients in a way that minimizes error.

Given the prevalence of linear models and the tendency of scientists to turn tools into theories, are we, as social scientists, at risk of seeing the world through the lens of a linear model? I would argue that in my area of specialization, the psychology of decision-making, this is the case. In particular, I think that scholars of decision-making gravitate towards models that *weigh and combine* all available evidence, and towards models that *optimize* based on the information given, much like multiple-regression models.

As a prescriptive norm, there are old and abundant examples of weighing and combining evidence being equated with good reason. Consider the image of blind justice holding a scale, or Darwin's deliberations about whether to marry (Darwin [1887] 1969: 232–3), psychological theories based in making trade-offs (e.g. Fishbein and Ajzen 1975), or the moral algebra Benjamin Franklin ([1772] 1987) offered as decision-making advice to a friend:

> [M]y Way is, to divide half a Sheet of Paper by a Line into two Columns, writing over the one Pro, and over the other Con. Then during three or four Days Consideration I put down under the different Heads short Hints of the different Motives that at different Times occur to me for or against the Measure. When I have thus got them all together in one View, I endeavor to estimate their respective Weights; and where I find two, one on each side, that seem equal, I strike them both out: If I find a Reason pro equal to some two Reasons con, I strike out the three. If I judge some two Reasons con equal to some three Reasons pro, I strike out the five; and thus proceeding I find at length where the Ballance lies; and if after a Day or two of farther Consideration nothing new that is of Importance occurs on either side, I come to a Determination accordingly. And tho' the Weight or Reasons cannot be taken with the Precision of Algebraic Quantities, yet when each is thus considered separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to make a rash Step; and in fact I have found great Advantage from this kind of Equation, in what may be called Moral or Prudential Algebra.

Moving from the prescriptive towards the descriptive, psychologists in recent theories of decision-making have commonly assumed that the mind works like a linear model, summing and weighing the evidence at hand. These range from multiple-regression-inspired 'lens model' applications (Hammond, Hursch, and Todd 1964) to the popular neural-network models of the 1980s and 1990s (Rumelhart, McClelland, and the PDP Research Group 1986), which weighed and combined evidence, and minimized error, much like multiple regression. In fact, simple neural networks are functionally equivalent to multiple regression. Under the view that minds make elaborate trade-offs, the modeling of decision behavior can be achieved in a straightforward way by simply fitting linear models to choice or ratings data. For example, one can present people with descriptions of apartments (rent, distance from city center, square footage, etc.) and ask them to rate the attractiveness of each apartment on a 100-point scale. Based on the ratings, one can simply run a multiple regression and interpret the resulting beta weights as the psychological 'decision weights' that the mind assigns to rent, square footage, and so on. This type of modeling is appealing because it is straightforward, and provides useful models that predict what kinds of apartments a person will prefer.

In sociology, certain rational-choice models embody the same spirit. For instance, an individual making a choice in accordance with subjective expected utility seeks all available options, considers the aspects (or possible outcomes) of each alternative, and assigns a weight to each. Stopping rules for gathering information, when present, can be optimization problems in their own right (e.g. 'stop at the point at which the cost of seeking out another alternative outweighs the benefits of having one'), and thus not address how a limited mind might make such a calculation. (For a discussion of the merits and limitation of rational-choice theory in sociology see Hechter and Kanazawa 1997 and Green and Shapiro 1994.)

Limitations of the rational, linear model view of cognition emerge in practice. In laboratory experiments people are usually given all the information they are supposed to consider in a decision, but outside the lab gathering information is expensive and getting all relevant information is impossible. Search costs often need to be paid with time or money. When a doctor needs to diagnose a patient who is on the verge of dying from an unknown cause, there are fifty tests she could run, but running all of them would bankrupt the patient, and could take more time than the patient has to live. A diagnosis based on limited information must be made. Similarly, search in memory takes time and people cut costs there as well—we don't pause in the grocery store, recalling our life experiences with each of the available brands of butter. In addition to information and memory-search costs, making trade-offs and optimal decisions invokes cognitive costs. In a split second a computer can run a regression and return a set of weights that not only allow for predictions of the outcome variable, but that take into account how each predictor variable might correlate with all the others. People, however, have tremendous difficulty keeping track of intercorrelations between cues (Armelius and Armelius 1974). Roger Shepard (1967) looked into the question of whether people integrate information like linear models do. While he noted that the perceptual system (i.e. the visual or auditory system) seems to carry out impressive feats of information integration, when it comes to higher-level cognition (i.e. reasoning about propositions) he found that people had trouble considering or combining even a few pieces of information. It has been noted that several schools of psychology have embraced the surprising thesis that higher-level cognition might work by simpler laws than lower-level cognition. Perhaps the eye and ear can weigh and integrate information better than the logical mind can (Todd and Gigerenzer 2000). The view of the mind as an optimal linear model is a view that ignores search costs and cognitive effort. It can at best be an 'as if' model.

Starting in the 1950s, this classical view of the mind was challenged by Herbert Simon, who proposed looking for models of 'bounded rationality' instead of classical rationality. Simon (1956) postulated that the time, information, and cognitive constraints of organisms necessitate that they 'satisfice' rather than optimize. Satisfice, a Scottish word blending 'satisfy' and 'suffice' means to settle for good-enough solutions, rather than for optimal ones. In the last 35 years the rational, optimizing view of the mind also came under attack by proponents of the successful heuristics-and-biases program (Kahneman, Slovic and Tversky 1982), who argued that much thinking is achieved through mental shortcuts or heuristics, which by their very nature do not gather and combine all available information. The heuristics-and-biases program has retained the normative kernel of the classical view, for instance that weighing and combining all available information is what one *ought* to do, but rejects the idea that people can stand up to these ideals. Explanations in this research program have invoked rather high-level accounts for how people make judgments. A few of the program's constructs, such as *availability* and *representativeness*, and *anchoring and adjustment*, have been used as explanations in thousands of articles in psychology. In the early 1990s two new research programs introduced more precise, testable models of decision-making heuristics. Payne, Bettman, and Johnson's *adaptive decision maker* program (1993) forwarded the view that decision makers draw from a toolbox of heuristic strategies, and introduced process-tracing techniques to identify which individuals were using which strategy when. Following their lead, Gerd Gigerenzer, myself, and present and past colleagues from the Adaptive Behavior and Cognition research group at the Max Planck Institute for Human Development in Berlin have forwarded a view, based on what I have termed 'fast and frugal' models of cognition. Where the *heuristics and biases* program was more focused on departures from rational norms, the *fast and frugal* program rejected these norms as normative, and asked instead how well simple heuristics will fare when making inferences in real-world environments. While the adaptive-decision-maker program concentrated on identifying strategy use, the fast and frugal program concentrated more on the accuracy/effort consequences of the strategy selected. Finally, while both the adaptive-decision-maker and the *rational choice* paradigm of sociology concentrate on individuals making choices to satisfy their preferences, the fast-and-frugal program has chiefly looked at inferences; that is, minds making guesses about the world (and other minds).[2]

The goal of the fast-and-frugal program is to design and test computational models of heuristics that are (a) ecologically rational (i.e. they exploit structures of information in the environment), (b) founded in evolved psychological capacities such as memory and the perceptual system, (c) fast, frugal, and simple enough to operate effectively when time, knowledge, and computational might are limited, (d) precise enough to be modeled computationally, and (e) powerful enough to model both good and poor reasoning (Goldstein and Gigerenzer 2002). As with Simon's notion of satisficing, these models have *stopping rules* to limit search for information. As opposed to classical models, they do not weigh and combine information. While sacrificing the generality and intuitive appeal of linear models, fast-and-frugal models aim to offer robustness, transparency, and psychological plausibility.

In this chapter I describe the two heuristics that started the fast-and-frugal research program—the Recognition Heuristic and the Take-the-best Heuristic—

and then move on to describe how fast-and-frugal heuristics are being employed as models not just of individual decision-making, but of group decision-making as well. The problems that have been analyzed with the simple heuristics have been varied, including models that predict homelessness, high-school dropout rates, parental investment, obesity, attractiveness, mortality, house prices, rent, professor's salaries, fuel consumption, car accidents, pollution levels, mate search, stock-market returns, nonverbal social interactions, and beyond (Gigerenzer, Todd, and the ABC Research Group 1999). One chapter is not enough space to cover all aspects of judgment and decision-making, so the focus will be on the topic of inference.

7.1. Recognition-based inference

Imagine that you have made it to the final round of 'Who Wants to Be a Millionaire'. You will either leave the show with nothing or with a million dollars, based on your answer to the following question: 'Which city has a larger population: Milwaukee or Detroit?'.

As it turns out, we put this very question to two groups (though not for a million dollars). One was from the University of Chicago and the other from Munich, Germany. Which group was more accurate?

Naturally, more Americans than Germans should know that Detroit is the much larger city. Not only are University of Chicago students some of the best and brightest, but Chicago is rather close to both Milwaukee and Detroit, so first-hand experience and memory should prove useful. The Germans, on the other hand, knew little about American cities—in fact one-third of them had never even heard of Milwaukee before. Who did better? A pitiful 40 percent of the Americans (slightly worse than chance) answered correctly. However, the vast majority of the Germans (about 90 percent) got it right.

In a variation on this experiment we quizzed American students on the twenty-two largest cities in both the USA and Germany (Goldstein and Gigerenzer 2002). We were surprised again: Americans did slightly better with foreign cities (73 percent correct) than with domestic ones (71 percent correct). We refer to these situations, in which lesser states of knowledge can outperform greater states, as *less-is-more effects.*

How can people who clearly know less about a given domain nonetheless do better than those who know more? Consider the following thought-experiment. Three American siblings sit down to take a test on the hundred largest cities in Germany. The test consists of pairs of cities, and in each pair the larger city must be identified. Assume the youngest sibling has never even heard of Germany before, and recognizes none of the cities on the quiz. The eldest sibling recognizes them all. The middle sibling recognizes half of the hundred cities from what she has overheard. The cities she recognizes are larger than those she doesn't recognize in 80 percent of paired comparisons (a realistic assumption, as it turns out). Now assume (i) that the children know nothing about the cities beyond recognizing their names, and (ii) that they attack the test using a strategy called the *recognition heuristic*:

> Recognition heuristic: If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion.

How will the three children perform at the task? Since the recognition heuristic is only applicable when exactly one object is recognized, the youngest and eldest child will not be able to use it. These children will have to guess on every question and thus score 50 percent correct. (Why will the eldest child have to guess? Recall, the children know nothing about the cities beyond recognizing their names.)

For the middle child, who recognizes half of the objects, the questions will fall into three categories. In one quarter of the questions she will recognize neither city and have to guess. In another quarter of the questions she will recognize both cities and also have to guess. However, in the remaining half of the cities she will recognize one city but not the other. For these items she will pick the city she recognizes and score 80 percent correct. Achieving 80 percent correct on half the items and 50 percent correct (by guessing) on the other half, this child will attain 65 percent correct overall, counterintuitively outperforming even her elder sibling. The bottommost curve in Figure 7.1 shows how the three siblings, and all intermediate knowledge states, would perform on the quiz.

The recognition heuristic lends itself to mathematical modeling. If an individual uses the rule, the proportion correct on a paired comparison task is

$$2\left(\frac{n}{N}\right)\left(\frac{N-n}{N-1}\right)\alpha+\left(\frac{N-n}{N}\right)\left(\frac{N-n-1}{N-1}\right)\frac{1}{2}+\left(\frac{n}{N}\right)\left(\frac{n-1}{N-1}\right)\beta$$

where N is the number of objects in a reference class (for example 100 cities), n is the number of recognized objects (for example 50 cities), α is the probability of getting a correct answer when only one object is recognized, β is the probability of getting a correct answer when both objects are recognized, and 0.5 is the probability of getting a correct answer when guessing. If α and β remain constant as n varies, and $\alpha > \beta$, this function takes on an inverted-U shape, predicting that intermediate states of recognition knowledge can outperform more complete states. Figure 7.1 plots this function for 3 levels of β (0.5, 0.6, 0.7). Since these levels of β are less than α (which is 0.8 in the figure), the resulting curves achieve maximum accuracy when less than all the objects are recognized.

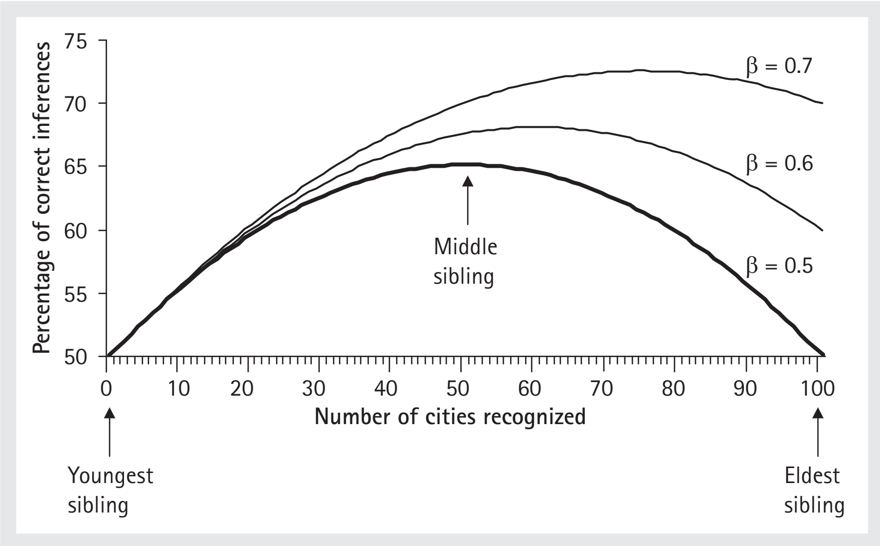

Fig. 7.1. Three siblings' responses to a test on the hundred largest cities in Germany. The accuracy of the three siblings from the thought-experiment, and of all intermediate knowledge states, is plotted on the bottom curve. To see what happens when we relax the assumption that the siblings know nothing about the cities beyond their names, we plot two lines depicting the case in which they have a 60 percent or 70 percent (β = 0.6 or 0.7) chance of making a correct inference when both objects are recognized. All three curves show how a less-complete state of recognition knowledge can, under the stated assumptions, lead to more correct inferences than a greater state of knowledge

The recognition heuristic is a simple model of two-alternative choice. It does not weigh or combine information. It limits search through a simple stopping rule ('Stop searching for cues if only one object is recognized') and it limits cognitive costs by having a simple choice rule ('Choose the alternative favored by recognition'). Despite its simplicity, in the interaction with natural environments it can make accurate inferences, predict choice behavior, and explain counterintuitive phenomena like the less-is-more effect.

Precisely defined heuristics can easily be coded into computer simulations to test hypotheses. We have used simulations to check whether the less-is-more effect derived above would arise when α and β are not fixed, but are allowed to vary in a real-world environment. The simulation learned to recognize German cities in the order of how well known they are. In addition to just learning to recognize cities, we ran conditions in which 0, 1, 2, or 9 pieces of information were learned in addition to city names. An example of a cue was the *soccer-team cue*: the simulation learned whether each city has a soccer team in the major league (an excellent predictor of population). After learning about each city, the simulated participant was quizzed on all pairs of cities, guessing when neither city was recognized and applying the recognition heuristic when exactly one city was recognized. When both cities were recognized, further cue knowledge, if any was available, was used to make an inference by means of another heuristic (named take-the-best, which we will look at later). Unlike the eldest sibling, who had to guess when both objects were recognized, our simulation could make a more informed inference in these cases. Would this cause the less-is-more effect to disappear? When the simulation had no knowledge beyond recognition, the less-is-more effect persisted. However, unlike the smooth curve in Figure 7.1, here we saw a jagged curve, but with the same inverted-U shape, reflecting the variation in α one might find when learning about cities in the real world. With one piece of information learned in addition to recognition, the simulation scored about 62 percent correct when all cities were recognized. However, when only three-quarters of the cities were recognized, the score was about 70 percent correct: a less-is-more effect. A similar pattern held when two or more pieces of information were available. Finally, with nine cues, the less-is-more effect flattened out.

As opposed to an all-purpose tool like a linear model, the recognition heuristic, or any simple heuristic, is domain specific. Instead of being unboundedly rational, it is *ecologically rational*, with respect to some environments but not others. There are domains in which the recognition heuristic will not work; for instance, to infer a US city's distance from Chicago. And there are domains in which people will not apply it. For instance, Oppenheimer (2003) presented people with pairs consisting of fictional cities with foreign-sounding names (e.g. 'Heingjing') compared to nearby small towns and found that participants chose the fictional, unrecognized cities 50 to 80 percent of the time. People may suspend use of the recognition heuristic when they believe test items constitute a biased sample or are drawn from an unknown reference class (e.g. the conjunction of neighboring towns and foreign cities) for which the validity of recognition is unknown. People might also not use the recognition heuristic when they have knowledge of the target variable (population), such as when they know that the town has less than 10,000 people. They might suspend use of the heuristic when they know that they recognize something from an artificial or uninformative context, such as from a lab experiment. The recognition heuristic is a model that lowers information-search costs, accordingly people might not use it when making inferences from information given to them on paper. In addition, the recognition heuristic is certainly not something that all people use in all situations. In every experimental test there are some participants whose decisions are best fit by recognition even in the face of conflicting cue information, and those who switch strategies in such cases. The *adaptive toolbox* view of cognition is one in which people learn, select, and switch between simple strategies, as opposed to update parameters in general-purpose models.

We have seen how simple heuristics can enable the specification of precise models and testing via analysis and computer simulation. To test models of behavior, experimental data can be compared to the heuristics' predictions. For example, to see how well the recognition heuristic predicts inferences, one can ask people which objects they recognize, give them paired-comparison questions, and then compute how often they choose recognized over unrecognized objects. Our studies on foreign cities found that despite having no free parameters, the recognition heuristic predicted choice in 90 to 100 percent of applicable cases. Other investigations have looked at the validity and adherence rates of the recognition heuristic for domains including predicting which athlete or team will win a competition, which party will win an election, which musician will sell more records, which stock will show higher growth, and beyond (Snook and Cullen 2006; Pachur and Biele 2007; Scheibehenne and Bröder 2007; Serwe and Frings, 2006).

Herbert Simon made the following observation of an ant walking on the beach:

> Viewed as a geometric figure, the ant's path is irregular, complex, hard to describe. But its complexity is really a complexity in the surface of the beach, not a complexity in the ant . . . The ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself. ([1969] 1996: 51–2)

Simple heuristics are able to make complex predictions by exploiting complexity that is inherent in the environment, such as the correlations between recognition and objects in the environment, and how these change between people, between domains, or over time. We turn now to show how a similar approach can be taken to a somewhat more difficult problem: making inferences with knowledge beyond recognition.

7.2. Knowledge-based inference

What is the simplest way to make smart choices between recognized alternatives? Sticking with our example domain, assume someone's knowledge about four German cities can be represented as in Table 7.1:

The table represents subjective knowledge. A plus in the top row means that the person recognizes that city, while a minus means they do not. We use the term 'cue' to mean a piece of information, and the cues in this table take on three values: plus, minus, or question mark. For the recognized cities, a plus means that the person believes the city has a feature; for instance, this person believes that city A was once an exposition site and believes that city B has a major-league soccer team. A minus represents the belief that a feature is lacking, so this person believes that city A is not

Table 7.1. A subjective knowledge matrix

	City			
	A	B	C	D
Recognition	+	+	+	–
Cue 1: Exposition site	+	–	?	?
Cue 2: Soccer team	?	+	?	?
Cue 3: Intercity train	–	+	+	?
Cue 4: State capital	?	–	–	?
Cue 5: University	?	?	–	?

on the intercity train line, city B is not a State capital, and city C lacks a university. If a person does not know whether a city has a particular feature, it is represented by a question mark. Note that the unrecognized city has question marks for all features, which is an assumption we made in our simulations, though one could think of situations in which a person could infer or deduce the cue values of an unrecognized object.

How might a mind or machine make an inference from basic knowledge such as this? A model could embrace the Enlightenment notion of classical rationality, and assume that when making a decision all information will be searched for, weighed, and combined optimally. On the other hand, a modeler could ask what the simplest possible mechanism might look like. We took the latter approach in designing what we have called the minimalist heuristic and the take-the-best heuristic.

7.2.1. Minimalist heuristic

The minimalist heuristic proceeds as follows:

Step 0. If applicable, use the recognition heuristic; that is, if only one object is recognized, predict that it has the higher value on the criterion. If neither is recognized, then guess. If both are recognized, go on to Step 1.

Step 1. Random search: Draw a cue randomly (without replacement) and look up the cue values of the two objects.

Step 2. Stopping rule: If one object has a positive cue value (+) and the other does not (i.e. either – or ?) then stop search. Otherwise go back to Step 1 and search for another cue. If no further cue is found, then guess.

Step 3. Decision rule: Predict that the object with the positive cue value has the higher value on the criterion.

For instance, when deciding whether City B or City C is larger, the minimalist heuristic might first try cue 4 at random. Since it doesn't discriminate between the two cities (neither is a state capital), another cue is drawn, say cue 3. This cue also doesn't discriminate, and so now cue 2 might be drawn. Since City B has a professional soccer team and City C may or may not, the stopping rule is satisfied, and City B is chosen.

7.2.2. Take-the-best heuristic

The minimalist heuristic certainly is simple. However, it is safe to assume that organisms know that not all cues are created equally. Even pigeons can learn the relationship between a cue and a criterion, so a model that tries cues in a particular order may be warranted. But which order? The take-the-best heuristic orders cues by *subjective cue validity*, which is a subjective estimate of *objective cue validity*. Objective cue validity is a feature of the environment. It is the relative frequency with which a cue correctly predicts the criterion defined with respect to the reference class. For instance, in the set of German cities we studied (Gigerenzer and Goldstein 1996), when inspecting all pairs of cities in which one has a soccer team and the other does not, the cities with teams have a larger population 87 percent of the time. Thus, the objective validity of the soccer team cue is 0.87. Take-the-best only differs from minimalist in one way. Instead of step 1 being random search, in take-the-best we have:

Step 1. Ordered search: Choose the cue with the highest subjective validity that has not been tried before. Look up the cue values of the two objects.

Heuristics are composed of building blocks. By changing one step, a very different model can emerge. Minds, we hypothesize, can alter the building blocks of cognitive strategies to match new task environments. These are the building blocks of the take-the-best heuristic:

- Search rule: Look up cues in order of validity
- Stopping rule: Stop search after the first cue discriminates between alternatives
- Decision rule: Choose the alternative that this cue favors

There are many contexts in which a mind might need rules for searching, stopping the search, and for choosing. While these categories of rules are rather general, the specific instantiation of the rules (e.g. random versus ordered search) may vary from situation to situation. In some domains, where abundant reasons are needed to justify a decision, instead of using take-the-best, a decision maker might search for cues until two are found that point to the same alternative. This modifies the stopping and decision rules, but still provides a precise, plausible model.

Why are minimalist and take-the-best fast and frugal? Search costs are reduced by only looking for as much information as is needed to discriminate between the alternatives. Cognitive costs are reduced because neither model deals with the correlations between the cues, neither tries to weigh and combine predictors. Both have a very simple choice rule: choose the object to which the first discriminating cue points.

Gerd Gigerenzer and I compared the performance of take-the-best and minimalist to that of strategies that use more information, carrying out a 'horse-race' simulation in which strategies competed (Gigerenzer and Goldstein 1996). The contestants included:

1. The minimalist heuristic;
2. The take-the-best heuristic;
3. Dawes's rule: Give each object a score, which is the number of positive cues minus the number of negative cues. Choose the object with the higher score;
4. Franklin's rule: Give each object a score, which is the sum of the cue validities of the positive cues minus the sum of the cue validities of the negative cues. Choose the object with the higher score;
5. Multiple regression: Fit a multiple-regression model to the matrix of cues and criterion values, and use the resulting linear model to arrive at a score for each object based on its cue values. Choose the object with the higher score.

The environment used was a set of German cities with more than 100,000 inhabitants (83 cities at the time). Exhaustive pairs of objects were drawn from this reference class and each competing strategy attempted to infer the larger object in each pair, based on the cue values. Missing knowledge was simulated in two ways. First, there were states of limited recognition knowledge in which 0 to 83 cities were recognized. As in the above thought-experiment, and consistent with our empirical estimates, the recognized cities were larger than the unrecognized cities in 80 percent of all possible pairs. The second dimension of missing knowledge was missing knowledge about cue values. From the complete matrix, we randomly replaced cue values with missing values, such that 0, 10, 20, 50, 75, or 100 percent of the original cue values remained. At each state of limited recognition knowledge and limited cue knowledge the results of many trials were averaged so that the subsets of recognized cities and missing cues would be different from trial to trial.

How well did the strategies perform? Before looking at their accuracy, let us consider the disadvantage the two fast-and-frugal heuristics faced. Unlike the linear models (contestants 3–5), the heuristics used limited information to make inferences. The linear models looked up all 10 cues per object, while take-the-best looked up only 3 on average, and minimalist only 2.8. How well can a heuristic do if it uses less than one-third of the available information? Table 7.2 shows the accuracy of all the competitors, averaged across all states of missing knowledge. Despite having access to all the cues, Dawes's rule and Franklin's rule were actually outperformed by the fast-and-frugal heuristics. Multiple regression, which has the benefit of all the information in addition to optimized weights, did as well as take-the-best.

Table 7.2. Two fast-and-frugal heuristics versus three linear models

Strategy	Frugality (Number of cues looked up)	Knowledge about Cues	Accuracy (% correct)
Take-the-best heuristic	3	Rank order	65.8
Multiple regression	10	Beta weights	65.7
Franklin's rule	10	Cue validities	62.3
Dawes's rule	10	Direction	62.1
Minimalist heuristic	2.8	Direction	64.7

Note: The number of cue values looked up is calculated per object in the comparison. The 'direction' of a cue is its sign: positive or negative.

The above contest looked at only one environment, and was a fitting, rather than a prediction, task. That is, the models got to train themselves on the same data on which they were later tested. If scientific models are to help us predict the future and generalize to new situations, they should be judged on their ability to predict new data, not merely on how well they fit available data. Jean Czerlinski, Gerd Gigerenzer, and I increased the number of environments from 1 to 20, so that, instead of looking only at German cities, we drew upon statistical databases and made inferences about a diverse set of topics, ranging from sociology to psychology to biology (Gigerenzer, Todd, and the ABC Research Group 1999). Environments varied in size from 11 objects to 395 objects and from 3 cues to 18 cues. The predictive quality of the strategies was tested using cross-validation: each environment was split in half randomly, and models were trained on one half, and tested the other. (A thousand random splits were carried out for each environment.) 'Training' has different meanings for different models, as some estimate more parameters than others. Training for multiple regression means determining a set of weights. For take-the-best it means determining a cue order. For Dawes's rule and minimalist, training only involves figuring out the direction of the cues (for example, figuring out whether having a soccer team correlates with being a large or small city). In contrast to the previous simulations, all objects were recognized. If the good performance of the heuristics in the previous simulation was overly due to the recognition heuristic, in these twenty environments there would be no such advantage.

How did the various strategies perform across more data sets and cross-validation? Looking at frugality, the heuristics maintained their advantage. Multiple regression and Dawes's rule each looked up 7.7 cue values on average while minimalist and take-the-best looked up 2.2 and 2.4 respectively. Figure 7.2 shows accuracy in both fitting and prediction. In fitting, multiple regression comes in first at 77 percent correct. Take-the-best again beat Dawes's rule. When moving from fitting to prediction all strategies suffered, but regression's accuracy decreased the

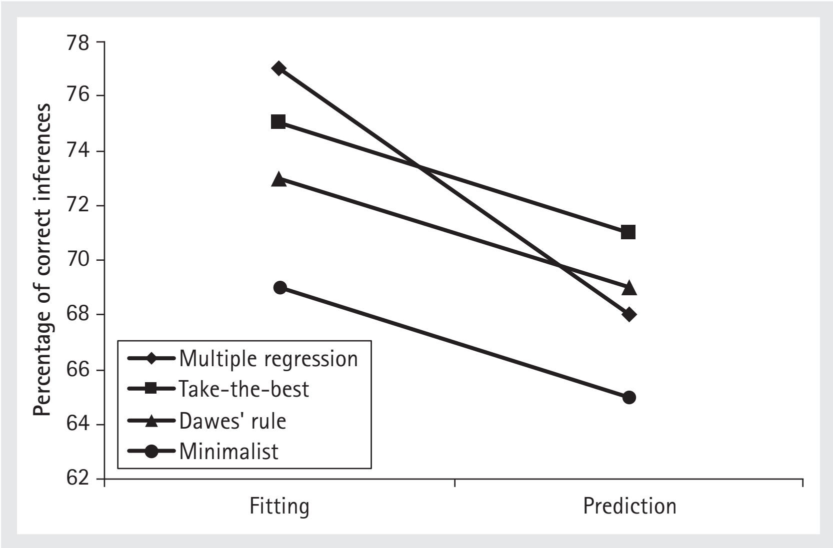

Fig. 7.2. Accuracy of the decision rules tested by Czerlinski, Goldstein, and Gigerenzer (1999) in fitting (training and testing on the same data) versus prediction (training on half the data and testing on the other half). Because of the tendency of models with more free parameters to overfit the training data, multiple regression loses the most accuracy when moving from fitting to prediction

most. As a result, take-the-best won the predictive competition, Dawes's rule came in second, and multiple regression dropped to third.

How can a strategy that uses partial information (like take-the-best), and a strategy that gives equal weight to all cues (like Dawes's rule), outpredict multiple regression, which determines its weights in an error-minimizing way? One reason is that more complex models, especially models with many free parameters, run the risk of overfitting training data. The training set can be thought of as having both structure and noise. A good model treats the structure as structure and regards the noise as noise, but an overparameterized model will treat the noise as part of the structure. When facing a prediction task, in which the structure will be the same but the random noise will be different, this overparameterized model can make worse predictions than a simpler variant. When a model performs well in generalizing from test sets to training sets, it is called robust. Robyn Dawes, for whom Dawes's rule is named, published an excellent paper in 1979 entitled 'The Robust Beauty of Improper Linear Models in Decision Making'. The simple 'unit' weights (+1 and −1) in Dawes's model allow for robust predictions because noise in the training data usually does not affect the weight. The worst that could happen is that the weight would change sign (i.e. go from + to − or vice versa) if the training data were very different than the test data. Take-the-best uses a cue ordering instead of

978–0–19–921536–2 Hedstrom-Main-drv (Typeset by SPi, Chennai) 155 of 734 March 7, 2009 22:49

unit weights or beta weights, and we have seen that it often relies on only a few cues. These results suggest that the cues which are highly valid in the training set tend also to be highly valid in the test set.

Much of the research in the field of judgment and decision making has historically involved decision-making at the individual level. In the interests of this volume, I'll turn now to recent research in which simple heuristics are being used to model social behavior.

7.3. Social heuristics

Research on social heuristics tests the assumption that models of group decision-making can be created with the same building blocks (stopping rules, choice rules, etc.) that individual heuristics employ. Since the building blocks were designed to be as simple as possible, this approach to modeling may allow us to discover the minimal level of complexity needed to model complicated social behaviors.

7.3.1. The recognition heuristic in groups

Many important decisions, for instance who gets the Nobel Prize, are made by groups. It has been observed that people seem to win prizes for winning prizes, that is the rich get richer, which has been called 'Matthew effect' in science (Merton 1968; see also the chapters by Podolny and Lynn and Hedström and Udehn in this volume). How might this arise? Consider that many prize committees are interdisciplinary, and since the members come from different fields, they may not be familiar with the same candidates. In fact, they may not even recognize the names of many potential candidates. If the group is guided by an invisible heuristic that says 'only give the prize to a candidate whose name is recognized by all the committee members', it may lead to awarding the prize to an already famous person, since any promising candidate who is unanimously recognized by an interdisciplinary group has a good chance of already being famous.

Could a group's inferences be guided by a simple recognition-based heuristic? Torsten Reimer and Konstantinos Katsikopoulos (2004) looked at what inferences groups make when some of the members only recognize some of the alternatives. In the first part of their experiment, German students were given a recognition test on American cities, and asked which city is larger in all pairs of recognized cities. In the second session participants were placed into groups of three and given pairs of cities, now with the opportunity to discuss before arriving at a group decision of which is larger. Since the experimenters knew precisely which cities each member

Table 7.3. Three individuals' responses to two cities in a group decision-making task

Member	Recognize Milwaukee	Recognize Detroit	Belief
1	Yes	Yes	Milwaukee is larger
2	Yes	Yes	Milwaukee is larger
3	No	Yes	–

recognized, and what each member believed about the relative populations of the recognized cities, hypotheses concerning group recognition could be tested.

Consider the case in Table 7.3. The question is to infer whether Milwaukee or Detroit has a larger population. Members 1 and 2 recognize both cities, and believe that Milwaukee is larger. Member 3, however, only recognizes Detroit. What will the group decide? The opinions of numbers 1 and 2 ought to carry considerable weight, as recognizing both cities implies impressive knowledge of US geography. If the group were to vote based on their previously expressed beliefs, the majority of members would vote for Milwaukee. However, in these situations, the experimental groups *rejected* the majority belief about 60 percent of the time. Partial ignorance (incomplete recognition) by one group member spoke louder than majority recognition.

If we change Table 7.3 such that member 2 recognizes neither city, a similar situation occurs. This is a case in which one person, who knows more about the objects, has expressed a belief which goes against the prediction of a recognition-based heuristic. Much like in the previous example, when these cases were analyzed the group chose the object recognized by the less-knowledgeable member about 60 percent of the time. When two group members had incomplete recognition, and the third believed that the city not recognized by the other two is larger, the group chose the object recognized by those with partial recognition about 75 percent of the time. A lack of knowledge can speak louder than expressed beliefs in certain instances of group decision-making.

7.4. Knowledge aggregation in groups

Anyone who has served on committees can attest that groups do not always make the best decisions. One example problem, which has captured the attention of social psychologists for years, is the *hidden profile* task. Suppose a group of four members must choose which of two job candidates to hire: Candidate A or Candidate B. There are facts (henceforth *cues*) that speak in favor of each candidate, numbered 1 to 12, and shown above the line in Figure 7.3. Examples of cues might be proficient typing ability, the ability to speak French, a university degree, and so on. Cues 1 through 4 are common knowledge; that is, all four group members know these facts. In addition, Cues 1 through 4 support candidate A. The remaining cues 5–12 are not shared (i.e. each is known only by one person), and support Candidate B.

Numerous experiments have been conducted in which groups of participants are assembled, given shared and unshared cues by the experimenter, and asked to choose a candidate. What might the groups decide? If the group were to pool all available information about candidates A and B, the collective knowledge of the group would be as represented below the line in Figure 3. Candidate A would have four supporting reasons, while Candidate B would have eight, suggesting, *ceteris paribus*, that B should be chosen. Because the candidates can only be seen

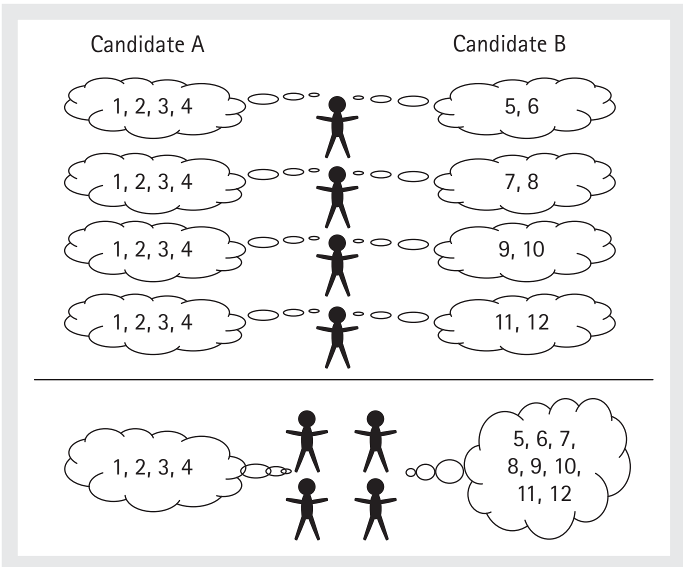

Fig. 7.3. Distribution of knowledge in a 'hidden profile' scenario. Cues (numbers 1 through 12) concerning two job candidates (A and B) are distributed among four group members who must decide to hire one candidate. Each individual has knowledge that suggests Candidate A is better than Candidate B, but if they were to pool their knowledge as a group—the situation depicted below the line at bottom—they would see there are more cues supporting Candidate B

in this light after pooling information, the full representation is called a 'hidden profile'.

What if the group members do not pool their information? From the point of view of each committee member, Candidate A has four pros while Candidate B only has two, so each individual might arrive at the opinion that A is better. Were the group to vote, one can see how Candidate A could be chosen over Candidate B. When such problems are given to groups of people in experiments, the typical result is that Candidate A is chosen. One characteristic finding is that groups tend to talk more about candidate A than B; that is, they focus more on shared than unshared information (Stasser and Titus, 1987; Wittenbaum and Stasser, 1996; Gigone and Hastie, 1997).

One explanation of groups' failure to detect hidden profiles is that they do not invest sufficient time and effort gathering and processing information. Is it true that extensive information-processing is needed to solve this problem? Are there simple strategies groups might employ to make fast, frugal, and accurate decisions? Torsten Reimer, Ulrich Hoffrage, and colleagues have extended the family of simple heuristics to address this task and found that a modified version of the take-the-best heuristic could effectively detect hidden profiles in computer simulation, suggesting that extensive information-processing is not necessary to solve the hidden-profile task. This result also poses new research questions. Why, if simple strategies exist, do real groups fail? How common are hidden profiles, anyway? When there is a hidden profile, how often is it the best choice? Reimer and Hoffrage (2005) found, when randomly generating scenarios (that is, creating simulated individuals who possess varying amounts of knowledge about the candidates) that only 456 in 300,000 environments contained a hidden profile—not very often, considering the amount of attention they have received in the literature. Detection failures could be significant if the hidden alternative (which groups typically do not choose) is almost always the best one. By running their simulations using a criterion of quality, Reimer and Hoffrage found that out of three alternatives, the hidden profile was the highest-valued one from 41 to 71 percent of the time, meaning that in a number of cases it could be wise *not* to choose it. In addition, they found that hidden profiles were most likely to arise when cues were low in validity or steeply decreasing in validities, and they found that hidden profiles were much more likely to be the best alternatives when cues were compensatory and high in validity. All these results are useful to predict when hidden profiles can be a cause for concern in applied settings.

Another example of heuristic methods for group problem solving comes from research by Mason, Jones, and Goldstone (2005). Using sophisticated software, laboratory participants sitting at computers were connected to one another through lattice, 'small world', full, and random network structures. Individuals were given a difficult problem of trying to find the maximum value of an unknown function by trial and error. They could submit guesses (numbers in the range of 1 to 100) to the computer and receive as feedback the value of the function for that guess, but with random noise added to make matters difficult. In addition, they could see the guesses and feedback of their immediate neighbors in the network. In an initial test the function to be maximized was unimodal, and perhaps not surprisingly people in fully connected networks got close to the maximum fastest. Since every player could see every other player's guesses, people likely imitated best guesses, leading all players to do their prospecting in an ultimately favorable region of the search space. However, a different picture emerged in a second condition in which the unknown function was not unimodal, but had a global maximum and two broad local maxima that could trap guessers. One might expect that, again, the greater amount of information available in the fully connected network would help its members. Surprisingly, in the trimodal problem the small-world network was the quickest to approach the global maximum. Individuals in the fully connected network were less exploratory and a strategy of imitation often led them into suboptimal local solutions. People in the small-world network were motivated to explore their local parts of the problem space, but since they were connected by 'shortcuts' to other locally working groups, they could be made aware of better solutions. Here we can see how the success of a simple heuristic like imitation depends not just on the structure of the information environment (unimodal versus multimodal), but on the social networks that exist between decision makers. We also observe an instance in which less information leads to better decisions through its effect on local exploratory behavior.

7.5. Learning cue orders socially

Heuristics like take-the-best search for cues in a particular order. In earlier work (Gigerenzer and Goldstein 1996) we did not put forward theories of how cue orders were learned. Our hunch at the time was that it happened through cultural transmission. Todd and Dieckmann (2005) have modeled the cue-order learning process, and run simulations in which decision makers try to learn on their own. They found that when decision makers learned in isolation, they arrived at good cue orders slowly. Even after feedback on a hundred paired comparisons, accuracy rose only modestly above that obtained with a random cue ordering. A similar result was found in the lab as well: Todd, Gigerenzer, and the ABC Research Group (in press) gave people feedback on a hundred items and noted they exhibited slow progress in learning cue orderings.

How might many minds work together to speed up individual learning? Garcia-Retamero, Takezawa, and Gigerenzer (2006) created a 'social learning' simulation that consisted of two alternating phases. In one phase individuals attempted to learn on their own, getting feedback from the environment. After a certain number of trials, individuals would come together in the other phase to share what they had learned. After this, they would go back to making inferences on their own with an (ideally) improved cue ordering and repeat the cycle. Figure 7.4 shows the situation in the Todd and Dieckmann simulation at the top, in which people learn as individuals and do not communicate, and that of the Garcia-Retamero, Takezawa, and Gigerenzer simulation, in which people alternate between individual and social learning phases, at the bottom.

In the social-learning simulation, individuals start out on their own (lower-left section of Fig. 7.4) making inferences with a random cue ordering. Supposing there are five cues called A, B, C, D, and E, this initial ordering might be D, C, A, E, B. After learning individually, each person might reorder the cues according to a specific criterion; for instance, the success rate of each cue when it was used. When coming together (lower-right section of Fig. 7.4), the members must decide on an ordering collectively. The individuals will each take this ordering back into the world (back to the lower-left of Fig. 7.4), modify it, and repeat.

As a group, there are many knowledge-aggregation rules that might be used to arrive at a new cue ordering for all to adopt. Several combination rules were tested in simulation, including majority voting, Condorcet voting, as well as averaging the cue-validity estimates of all members. These all involved combining information from all the members into the final decision. A very simple rule, which ignores the input of all but one member, is having everyone simply imitate the cue ordering of the individual that was most successful on his or her own. Though this rule was simplest and most frugal, it had the surprising result of exhibiting the fast learning rate and greatest accuracy, as Figure 7.5 shows. This social learning can be very rapid. One simulation found that if a group of 10 individuals meets just once (after each getting feedback on 50 paired comparisons), the accuracy is as high as that obtained when ordering cues according to their true ecological validities. Mechanisms like these help us see how potentially difficult aspects of applying individual heuristics (such as arriving at a cue ordering on one's own) can be plausibly modeled through simple social interaction. The success of imitation strategies, however, depends on the rational abilities of each agent to faithfully record, and the honesty of each agent to faithfully report, its individual success, as well as the ability of agents to adjust for the possibly biased or dishonest information coming from others. (For further discussion of the consequences of biased agents for simulations see the chapter by Macy and Flache, and for more on the issue of whom to trust see the chapters by Bohnet and by Cook and Gerbasi, all in this volume.)

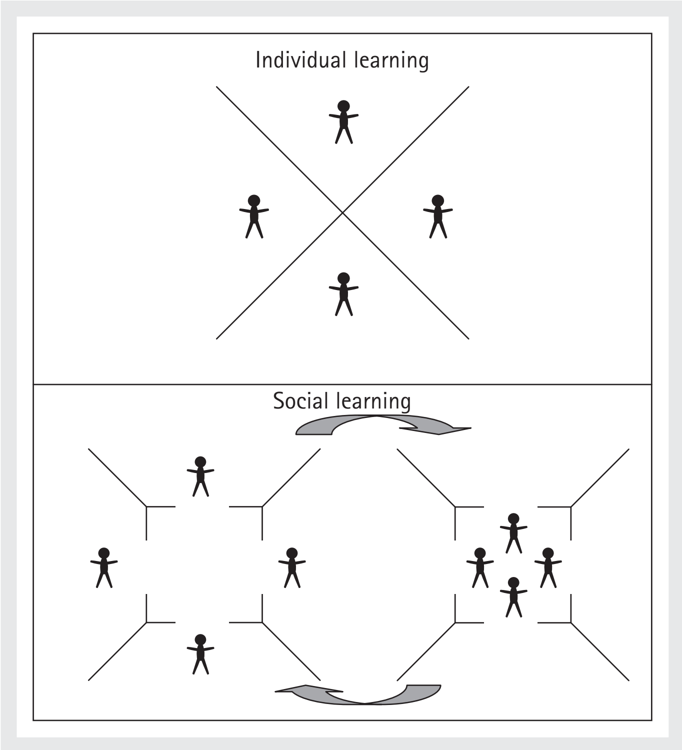

Fig. 7.4. Individual and social learning as modeled by Todd and Dieckmann (2005) and Garcia-Retamero, Takezawa, and Gigerenzer (2006). The objective is to learn a cue ordering for take-the-best. At top, in individual learning, decision makers get feedback from the environment only, but not from each other. In social learning, at bottom, the simulation alternates between an individual learning phase at left, and a social exchange phase at right. In social exchange, group members communicate the experiences they had during individual learning, in order to collectively arrive at new cue ordering as a group. Once a new ordering is found, all individuals test it out alone and the cycle repeats

Considerations

The fast-and-frugal approach has some limitations. In order to keep the definitions of heuristics precise, they may need to deal with operational definitions that can gloss over subtleties of psychological experience. The recognition heuristic, for instance, uses as an input a yes/no judgment concerning whether an object is

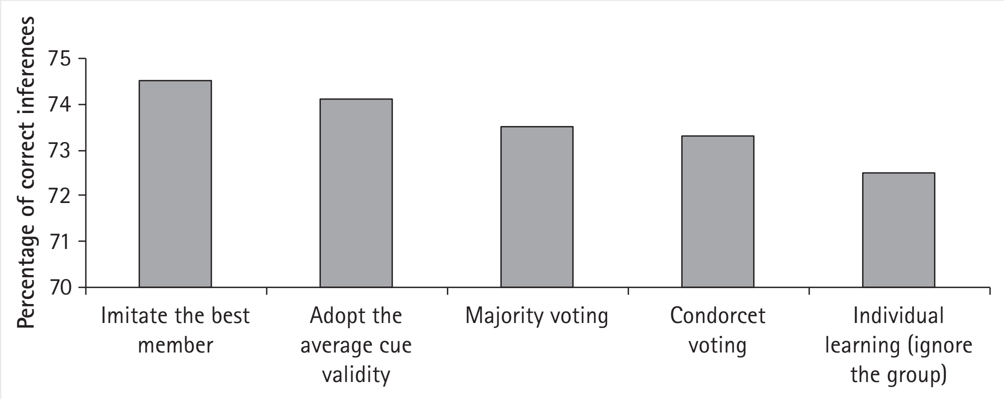

Fig. 7.5. Correct inferences under various social learning rules after a hundred paired comparisons (Adapted from Garcia-Retamero, Takezawa, and Gigerenzer 2006). A frugal strategy of ignoring all the input except that of the most successful member fared best. As benchmarks, the minimalist heuristic achieved 70% correct, and take-the-best (using the true ecological validities) scored 74.1% correct

recognized. This judgment, however, can be influenced by numerous moderating factors (Dougherty, Franco-Watkins, and Thomas 2008; Gigerenzer, Hoffrage, and Goldstein 2008). For instance, a slightly familiar name may be rated as recognized in the context of unfamiliar names, but unrecognized in the context of very famous names. These shifts in recognition judgments can lead to changes in inference. One needs to look beyond the simple recognition heuristic to model how recognition judgments change according to context. Similarly, contextual cues may cause a person to switch from using one heuristic to using another. A person might use the recognition heuristic for items that are naturally sampled from the environment, but suspend its use on perceived trick questions. Reinforcement learning can cause people to switch between heuristic strategies as well, and Rieskamp and Otto (2006) have recently put forward a promising theory of how people learn to do so.

It is tempting to criticize simple and noncompensatory heuristics because one can easily identify 'corner' cases in which they should make incorrect predictions. Take-the-best or the recognition heuristic may arrive at the wrong answer when the first cue points towards one alternative and all the other cues point to the other. Furthermore, in such cases simple heuristics may mispredict behavior as well. However, models should be evaluated not on the capacity to mispredict in theory, but on the number and consequence of mispredictions in realistic tests. Linear models mispredict, too; it is just harder for humans to notice a priori when and where they will break down. It is also difficult at first glance to appreciate that simpler models can be more robust—the results of Figure 7.2 come as a surprise to many.

For every heuristic there are kind and unkind environments. A heuristic that works well in the natural world, for instance, may fail in social situations in which players are trying to mislead each other. Perhaps more than general-purpose models, the accuracy and efficiency of heuristics are tightly linked to properties such as cue validities, the cost of information, cue intercorrelations, rates of change, likelihood of false information, local maxima, network structure, and more. Through computer simulation, mathematical analysis, and laboratory experimentation, the science of heuristics is getting a clearer picture of how the success of strategies depends on the structure of the environment (Katsikopoulos and Martignon 2006; Hogarth and Karelaia 2007; Todd, Gigerenzer, and the ABC Research Group, in preparation).

This chapter provides examples of how simple heuristics might account for complex behavior, how they can serve as robust models, and how they can make counterintuitive predictions. Perhaps the greatest benefit of this heuristic-based approach is its precision. Hedström and Swedberg (1998) have decried vagueness in social theory, noting a tendency 'to label, relabel, and describe rather than to explain', and similar critiques have been aimed at explanation by redescription in psychology (Gigerenzer 1996). When developing the models in the book *Simple Heuristics That Make Us Smart* (Gigerenzer, Todd, and the ABC Research Group 1999), our research group was guided by the maxim 'if you can't write it as a simple computer program, it's not a simple heuristic'. While coding something is not a high hurdle (anything can be written as a program with enough assumptions), and while not all computer programs clarify matters (e.g. complex neural networks which remain opaque even to their creators), the rule turned out to be valuable in practice. Not only did it lead to models that were more precise, but the maxim led to accidental discovery as well. For instance, while coding a variant of take-the-best called take-the-last (which tries cues in order which cue last discriminated), we were surprised when this simple mechanism won the competition by a full 10 percentage points (Gigerenzer and Goldstein 1996: 661). We had absentmindedly presented the questions in a systematic (as opposed to random) order that only this heuristic, and none of the more complex models like multiple regression, could exploit. Simon famously noted that rationality is bounded by a scissors whose two blades are 'the structure of task environments and the computational capacities of the actor' (1990: 7). Modeling simple heuristics and information environments in code leads one to notice the interaction of both blades.

There is a long tradition of viewing the mind as a generalized maker of tradeoffs, a prudent judge that weighs and combines all available information through mysterious optimizing processes. However, there is promise in coming at the problem from the opposite direction. By combining the simple building blocks of search rules, stopping rules, and choice rules, one can account for the complexities of behavior with models that are transparent, robust, and psychologically plausible.

Notes

1. The song by Ray Noble (1934), covered by Bing Crosby, Doris Day, Nat King Cole, Billie Holiday, and others.
2. All these approaches have, as one might imagine, addressed the questions that the others have posed. I arrive at these high-level characterizations by considering the kinds of data that each program has chosen to collect. For instance, heuristics-and-biases research often computes the proportion of people deviating from a normative answer, the adaptive-decision-maker program has collected ample process and preference data, and the fast-and-frugal approach has focused on algorithmic speed and accuracy.

References

Armelius, B., and Armelius, K. (1974), 'The Use of Redundancy in Multiple-cue Judgments: Data from a Suppressor-variable Task', *American Journal of Psychology*, 87: 385–92.

Coleman, J. S. (1986), 'Social Theory, Social Research, and a Theory of Action', *American Journal of Sociology,* 91 (6): 1309–35.

Czerlinski, J., Goldstein, D. G., and Gigerenzer, G. (1999), 'How Good Are Simple Heuristics?', in G. Gigerenzer, P. M. Todd, and the ABC Research Group (eds.), *Simple Heuristics that Make Us Smart* (New York: Oxford University Press).

Darwin, C. [1887] (1969), *The Autobiography of Charles Darwin, 1809–1882* (New York: Norton).

Dawes, R. M. (1979), 'The Robust Beauty of Improper Linear Models in Decision Making', *American Psychologist,* 34: 571–82.

Dougherty, M. R., Franco-Watkins, A. M., and Thomas, R. (2008), 'The Psychological Plausibility of the Theory of Probabilistic Mental Models and Fast and Frugal Heuristics', *Psychological Review,* 115: 199–211.

Edington, E. S. (1974), 'A New Tabulation of Statistical Procedures Used in APA Journals', *American Psychologist,* 29: 25–6.

Fishbein, M., and Ajzen, I. (1975), *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*, (Reading, Mass.: Addison-Wesley).

Franklin, Benjamin [1772] (1987), *Writings*, (New York: Library of America).

Garcia-Retamero, R., Takezawa, M., and Gigerenzer, G. (2006), 'How to Learn Good Cue Orders: When Social Learning Benefits Simple Heuristics', in R. Sun and N. Miyake (eds.), *Proceedings of the Twenty-eighth Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Erlbaum), 1352–8.

Gigerenzer, G. (1991), 'From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology', *Psychological Review,* 98: 254–67.

—— (1996), 'On Narrow Norms and Vague Heuristics: A Rebuttal to Kahneman and Tversky', *Psychological Review,* 103: 592–6.

—— and Goldstein, D. G. (1996), 'Mind As Computer: The Birth of a Metaphor', *Creativity Research Journal,* 9: 131–44.

—— Hoffrage, U., and Goldstein, D. G. (2008), 'Fast and Frugal Heuristics are Plausible Models of Cognition: Reply to Dougherty, Franco-Watkins, and Thomas (2008)', *Psychological Review,* 115 (1): 230–37.

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

Gigone, D., and Hastie, R. (1997), 'The Impact of Information on Small Group Choice', *Journal of Personality and Social Psychology*, 72: 132–40.

Goldstein, D. G., and Gigerenzer, G. (2002), 'Models of Ecological Rationality: The Recognition Heuristic', *Psychological Review,* 109: 75–90.

Green, D. P., and Shapiro, I. (1994), *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science* (New Haven, Conn.: Yale University Press).

Hammond, K. R., Hursch, C. J., and Todd, F. J. (1964), 'Analyzing the Components of Clinical Inference', *Psychological Review,* 71: 438–56.

Hechter, M., and Kanazawa, S. (1997), 'Sociological Rational Choice Theory', *Annual Review of Sociology*, 23: 191–214.

Hedström, P., and Swedberg, R. (1998), 'Social Mechanisms: An Introductory Essay', in Hedström, and Swedberg (eds.), *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge: Cambridge University Press), 1–31.

—— and Udéhn, L. (2009), 'Analytical Sociology and Theories of the Middle Range', in P. Hedström and P. Bearman (eds.), *The Oxford Handbook of Analytical Sociology* (Oxford: Oxford University Press).

Hogarth, R. M., and Karelaia, N. (2007), 'Heuristic and Linear Models of Judgment: Matching Rules and Environments', *Psychological Review*, 114 (3): 733–58.

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

Katsikopoulos, K. V., and Martignon, L. (2006), 'Naïve Heuristics for Paired Comparisons: Some Results on their Relative Accuracy', *Journal of Mathematical Psychology*, 50: 488–94.

Kelly, H. H. (1967), 'Attribution Theory in Social Psychology', in D. Levine (ed.), *Nebraska Symposium on Motivation*, (Lincoln, Nebr.: University of Nebraska Press), 192–238.

—— and Michaela, I. L. (1980), 'Attribution Theory and Research', *Annual Review of Psychology*, 31: 457–501.

Macy M., and Flache, A. (2009), 'Social Dynamics from the Bottom Up: Agent-based Models of Social Interaction', in P. Hedström and P. Bearman (eds.), *The Oxford Handbook of Analytical Sociology* (Oxford: Oxford University Press).

Mason, W., Jones, A. and Goldstone, R. L. (2005), 'Propagation of Innovations in Networked Groups', in M. Bucciarelli, B. G. Bara, and L. Barsalou (eds.), *Proceedings of the*

Twenty-Seventh Annual Conference of the Cognitive Science Society, 2005 (Mahwah, NJ: Erlbaum).

Merton, R. K. (1968), 'The Matthew Effect in Science', *Science*, 159 (3810): 56–63.

Oppenheimer, D. M. (2003), 'Not So Fast! (And Not So Frugal!): Rethinking the Recognition Henristic', *Cognition*, 90: B1–B9.

Pachur, T., and Biele, G. (2007), 'Forecasting From Ignorance: The Use and Usefulness of Recognition in Lay Predictions of Sports Events', *Acta Psychologica*, 125: 99–116.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993), *The Adaptive Decision Maker* (Cambridge: Cambridge University Press).

Podolny, J., and Lynn, F. (2009), 'Status', in P. Hedström and P. Bearman (eds.), *The Oxford Handbook of Analytical Sociology* (Oxford: Oxford University Press).

Reimer, T., and Hoffrage, U. (2005), 'Can Simple Group Heuristics Detect Hidden Profiles in Randomly Generated Environments?, *Swiss Journal of Psychology*, 64: 21–37.

—— and Katsikopoulos, K. (2004), 'The Use of Recognition in Group Decision-making', *Cognitive Science*, 28 (6): 1009–29.

Rieskamp, J., and Otto, P. E. (2006), 'SSL: A Theory of How People Learn to Select Strategies', *Journal of Experimental Psychology: General*, 135: 207–36.

Rumelhart, D., McClelland, J., and the PDP Research Group, (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, i. Foundations* (Cambridge, Mass.: MIT Press).

Scheibehenne, B., and Bröder, A. (2007), 'Predicting Wimbledon 2005 Tennis Results by Mere Player Name Recognition', *International Journal of Forecasting*, 23 (3): 415–26.

Schooler, L. J., and Hertwig, R. (2005), 'How Forgetting Aids Heuristic Inference', *Psychological Review*, 112: 610–28.

Serwe, S., and Frings, C. (2006), 'Who Will Win Wimbledon 2003? The Recognition Heuristic in Predicting Sports Events', *Journal of Behavioral Decision Making*, 19 (4): 321–32.

Shepard, R. N. (1967), 'On Subjectively Optimum Selections among Multi-attribute Alternatives', in W. Edwards and A. Tversky (eds.), *Decision Making* (Baltimore, Md.: Penguin), 257–83.

Simon, H. A. (1956), 'Rational Choice and the Structure of the Environment', *Psychological Review*, 63: 129–38.

—— [1969](1996), 'The Sciences of the Artificial', Karl Taylor Compton Lectures, MIT, 1968, 3rd edn. (Cambridge, Mass.: MIT Press).

—— (1979), 'Information-processing Models of Cognition', *Annual Review of Psychology*, 30: 363–96.

—— (1990), 'Invariants of Human Behavior', *Annual Review of Psychology*, 41: 1–19.

Snook, B., and Cullen, R. M. (2006), 'Recognizing National Hockey League Greatness with an Ignorance-based Heuristic', *Canadian Journal of Experimental Psychology*, 60: 33–43.

Stasser, G., and Titus, W. (1987), 'Effects of Information Load and Percentage of Common Information on the Dissemination of Unique Information during Group Discussion,' *Journal of Personality and Social Psychology*, 53: 81–93.

Todd, P. M., and Dieckemann, A. (2005), 'Heuristics for Ordering Cue Search in Decision-making', in L. K. Saul, Y. Weiss, and L. Bottou (eds.), *Advances in Neural Information Processing Systems*, xvii (Cambridge, Mass.: MIT Press), 1393–1400.

—— —— (in press). 'Simple Rules for Ordering Cues in One-reason Decision Making', in P. M. Todd and G. Gigerenzer, and the ABC Research Group (eds.), *Ecological Rationality: Intelligence in the World* (New York: Oxford University Press).

Todd, P. M., and Gigerenzer, G. (2000). 'Précis of Simple Heuristics that Make Us Smart', *Behavioral and Brain Sciences*, 23: 727–41.

—— —— and the ABC Research Group (in preparation), *Ecological Rationality: Intelligence in the World* (New York: Oxford University Press).

Wittenbaum, G. M., and Stasser, G. (1996), 'Management of Information in Small Groups', in J. L. Nye and A. B. Bower (eds.), *What's Social About Social Cognition?* (London: Sage), 3–28.