Simple Heuristics That Make Us Smart

Gerd Gigerenzer
Peter M. Todd
and the ABC Research Group

New York Oxford
Oxford University Press
1999

2

The Recognition Heuristic

How Ignorance Makes Us Smart

Daniel G. Goldstein
Gerd Gigerenzer

Human thought consists of first, a great capacity for recognition, and second, a capability for selective search.

Herbert A. Simon

On a country road in Scotland, MacGregor sees his old schoolmate MacAlister and calls out to him. MacAlister hesitates. He recognizes MacGregor's face, but has no idea of his name, where they had met before, or anything else. MacGregor expresses surprise that his old classmate would "tartle," that is, hesitate in getting beyond a sense of mere recognition. As this useful Scottish verb helps to demonstrate, recognition and recall memory can break apart. Sometimes this break can be permanent, as in the case of R.F.R., a 54-year-old policeman who developed amnesia so grave that he even had difficulty identifying his wife and mother in photographs (Warrington & McCarthy, 1988). It would seem that R.F.R. had lost his capacity for recognition. Had he? In an experiment, he was presented photographs of famous people and of strangers he had never seen before and was asked to point out the famous ones. He performed as though his memory were unimpaired. Even though he lacked the recall memory to name people (such as his mother) in photographs, he retained a normal recognition memory, and this allowed him to indicate the famous faces he had seen before.

Like R.F.R. and the tartling Scotsman, as we wander through a stream of sights, sounds, tastes, odors, and tactile impressions, we have little trouble knowing what we have encountered before, even when we cannot recall more information. Our sense of recognition is argued to constitute a specialized memory system that can be impaired independently of other memory capacities. For instance, elderly people suffering memory loss (Craik & McDowd, 1987; Schonfield & Robertson, 1966) and patients suffering certain kinds of brain damage (Schacter & Tulving, 1994; Squire et al., 1993) have problems saying what they know about an object, even where they have encountered it, but can act in a way that shows that they have encountered the object before. Similarly, laboratory research has demonstrated that recognition memory continues to encode information even in divided-attention learning tasks that are too distracting to allow more substantial memories to be formed (Jacoby et al., 1989). Mere recognition, this essentially binary feeling that we have or have not experienced something before, is a minimal state of knowledge. Why do minds encode it? What is mere recognition good for?

In this chapter, we introduce the simplest heuristic in this book, the *recognition heuristic*, which exploits the vast and efficient capacity of recognition to make inferences about unknown aspects of the world. The processes underlying face, voice, and name recognition are anything but simple, and are still far from understood in cognitive science. However, their output is available to us as a simple signal, recognition, which can be exploited by a very simple heuristic. The recognition heuristic is so frugal that it actually requires a beneficial lack of knowledge to work. In this chapter, we define the heuristic in the form of a simple rule, which allows us to study its performance by means of simulation and mathematical analysis. We show that, under certain conditions, it leads to the counterintuitive *less-is-more effect*, in which a lack of recognition can be beneficial for making inferences. We also illustrate how to measure recognition, which allows us to study experimentally whether people actually use the recognition heuristic.

The term "recognition" has been used in many contexts, so let us be clear about the way in which we shall use it. MacAlister steps onto a bus. The passengers may fall into three classes corresponding to the three columns in figure 2-1. There may be passengers he does not recognize, that is, whom he is sure he has never seen before, represented by the leftmost column. There may be passengers he merely recognizes, but whom he cannot identify or recall anything about (those that make him tartle), represented by the center column. Finally, there may be people he can recognize and also identify (what their profession is, for instance), represented by the rightmost column.

With the term "recognition," as the striped line in figure 2-1 shows, we divide the world into the novel (the leftmost column) and the previously experienced (the two rightmost columns). For instance, landmark recognition, which serves the adaptive function of helping an organism find its way home, is based on the simple binary distinction between the novel and the previously experienced. Mere recognition needs to be distinguished from degrees of knowledge and what is referred to as "familiarity," such as in theories that postulate that attitudes toward objects become more positively inclined with repeated exposure (e.g., Zajonc,

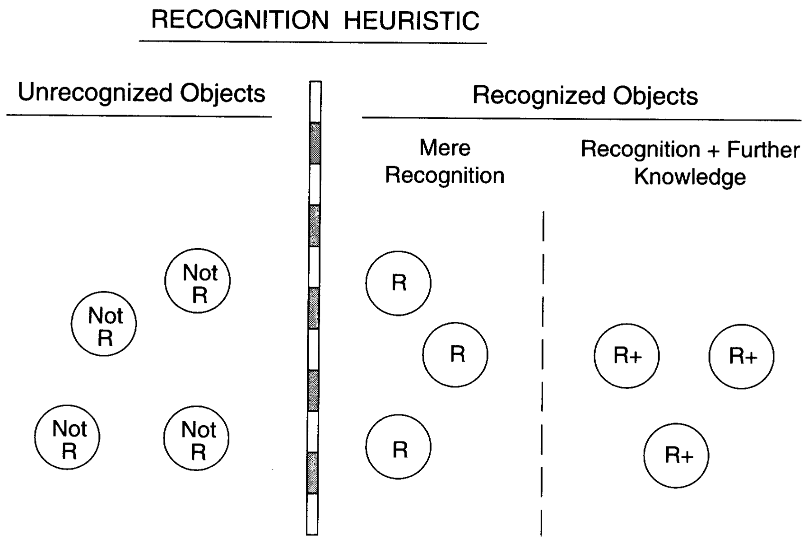

Figure 2-1: How the recognition heuristic applies to unrecognized, novel objects (Not R = not recognized), merely recognized objects (R), and objects about which something is known beyond recognition (R+). The distinction relevant for the recognition heuristic is that between unrecognized objects and everything else.

1968), and those that contend that the belief in an assertion increases with its repetition (e.g., Gigerenzer, 1984; Hasher et al., 1977). We will study heuristics that use knowledge beyond mere recognition beginning in chapter 4. Our use of the term "recognition" also needs to be distinguished from the very common usage that refers to a person's ability to verify whether an object was presented in a previous experimental session. Such studies often fail to touch upon the distinction between the novel and the previously experienced because the stimuli in these studies, mostly digits or common words, are *not* novel to the participant before the experiment. For example, "cat," a common word, would not be novel to someone before an experiment whereas "flink," a nonword, most probably would. In contrast, experiments that use never-before-seen photographs, as in the following examples, exemplify our sense of the word "recognition."

Recognition memory is vast, automatic, and save for déjà vu, reliable. Shepard (1967b) instructed participants to look through 612 pictures at their own pace and immediately afterward tested recognition memory with pairs of pictures, one previously presented and the other novel. Participants were able to recognize the previously presented pictures in 98.5% of all cases, on average. Standing (1973) increased the number of pictures (photographs and "striking" photographs preselected for their vividness) to 1,000 and limited the time of presentation to five seconds. In a test like Shepard's 48 hours later, participants were able to point to the previously presented picture 885 times (normal pictures) or 940 times (striking pictures). These figures become 770 and 880 after a correction for guessing. Standing then outdid himself by a factor of 10. In perhaps the most extensive recognition memory test ever performed, he presented people with 10,000 pairs of normal pictures from which participants chose correctly 8,300 times (6,600 with guessing correction). With respect to the performance with the "striking" pictures, Standing speculates, "if one million items could be presented under these conditions then 731,400 would be retained" (p. 210). Note that, while the retention percentage declines with the number of pictures presented, the absolute number of pictures recognized keeps increasing. We conjecture that the limits of recognition memory cannot be exceeded in a laboratory experiment, and perhaps not in the lifetime of a human being.

How to Benefit from Ignorance

The remarkable capacity for recognition in higher organisms is likely to have evolved for a number of adaptive functions. Consider the eating habits of wild rats, which exhibit strong *neophobia*, that is, a reluctance to eat foods they do not recognize (Barnett, 1963). This mechanism is adaptive in avoiding poisons: Every food a living rat has eaten has, necessarily, not killed it (Revusky & Bedarf, 1967). Norway rats prefer foods they recognize from having tasted them or from having smelled them on the breath of other rats (Galef, 1987; Galef et al., 1990). This heuristic for food choice is followed even if the rat whose breath is smelled happens to be sick at the time. That is, recognition dominates illness information. We will report later in this chapter on a related experiment with humans, in which recognition dominates conflicting information. Food choice in wild rats accords with the recognition heuristic, defined shortly.

In what follows, we describe the recognition heuristic and explore its inferential accuracy. We specify conditions under which this heuristic enables organisms with less knowledge to make more accurate inferences than organisms with more knowledge: a counterintuitive phenomenon we call the less-is-more effect. We start by introducing our "Drosophila" problem area—that is, an example that is well understood—for studying inference: geography.

Proper name recognition constitutes a specialized region in our cognitive system that can be impaired independently of other language skills (McKenna & Warrington, 1980; Semenza & Zettin, 1989; Semenza & Sgaramella, 1993). A person's knowledge of geography consists largely of proper names (those of cities, countries, mountains, and so on) and their assignment to real places on the earth. Geographical knowledge is always

incomplete, which makes it an ideal field of inquiry for studies of recognition. We will analyze recognition by means of computer simulation, mathematical analysis, and experimentation. In several demonstrations, we use a geographical topic about which our participants (students at the University of Chicago) had incomplete knowledge: cities in Germany. In particular, we dealt with the class of 83 German cities with more than 100,000 inhabitants. Our American participants recognized only about a quarter of these cities, and yet they were able to exploit this lack of recognition, as we will see.

The task we examine is a common one: selecting a subset of objects from a larger set. In this chapter, we focus on the case of choosing one object from two. This task, two-alternative choice, besides being a staple of experimental psychology, is an elementary case to which many problems of greater complexity (multiple choice, for instance) are reducible. An example of a two-alternative choice question is: "Which is the stronger currency: the pound or the markka?" Or, in the realm of geography: "Which city has a larger population: Munich or Dortmund?"

The Recognition Heuristic

Consider the task of inferring which of two objects has a higher value on some criterion (e.g., which is faster, higher, stronger). The recognition heuristic for such tasks is simply stated: *If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value.*

For instance, a person who has never heard of Dortmund but has heard of Munich would infer that Munich has the higher population, which happens to be correct. The recognition heuristic can *only* be applied when one of the two objects is not recognized, that is, under partial ignorance. Note that where recognition correlates negatively with the criterion, "higher" would be replaced with "lower" in the definition.

Recognition and the Structure of the Environment

The recognition heuristic is *domain-specific* in that it only works in environments where recognition is correlated with the criterion. How is the correlation between recognition and the criterion estimated? In some domains, the direction of this correlation can be genetically coded (as seems to be the case with the rat's inference that unrecognized food is suspect). In other domains, the direction of the correlation must be learned through experience. However, in cases of inference or prediction, the criterion is inaccessible to the organism. Though the criterion may be inaccessible, there are "mediators" in the environment that have the dual property of reflecting (but not revealing) the criterion and also being accessible to the senses, as figure 2-2 illustrates. A person may have no direct information

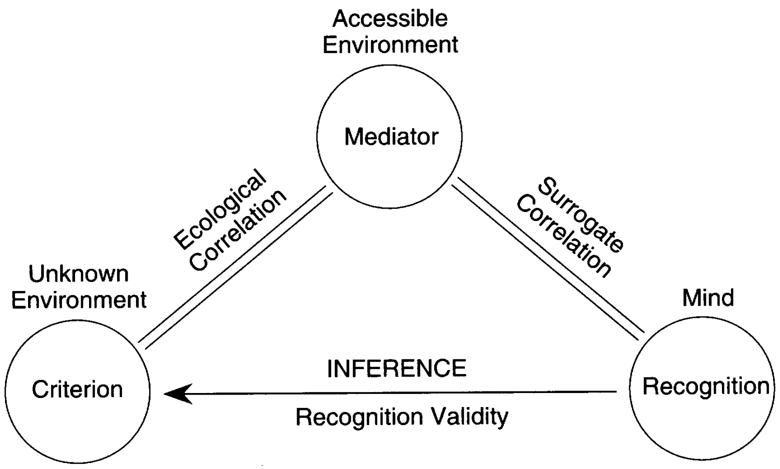

Figure 2-2: The ecological rationality of the recognition heuristic. The inaccessible criterion is reflected, but not revealed, by the mediator variable. The mediator influences the probability of recognition. The mind in turn uses recognition to infer the criterion.

about the endowments of universities, for example, as this information is not always accessible. However, the endowment of a university may be reflected in how often the university is mentioned in the newspaper. Since the newspaper is accessible, it is an example of a mediator. The more often a name occurs in the newspaper, the more likely it is that a person will recognize this name. For instance, Stanford University is more often mentioned in the national press than Miniscule State. Thanks to the mediator of the newspaper, a person can now make an inference about which of these two universities has a larger endowment. Three variables that describe the relationship between the criterion, mediator, and mind are the *recognition validity*, the *ecological correlation*, and the *surrogate correlation*.

The ecological correlation describes the relation between the criterion and the mediator. In the case of university endowments, the criterion is the endowment and the mediator variable is simply the number of times the university is mentioned in the paper (and not any information about its endowment). In the case of the Norway rats, the criterion is the toxicity of a food and the mediator variable could be the number of rats with that food on their breath (and not any other information concerning the health of these rats). The surrogate correlation is that between the mediator (which acts as a surrogate for the inaccessible criterion) and the contents of recognition memory. In our university example, the surrogate correlation is the number of times names are mentioned in the newspaper correlated against recognition of these names. Surrogate correlations can be measured against the recognition memory of one person (in which case the data will be binary), or against the collective recognition of a group, which we will demonstrate later.

The strength of the relationship between recognition and the criterion is the recognition validity, which we define as the proportion of times a recognized object has a higher criterion value than an unrecognized object in a given reference class. The recognition validity α is thus:

$$\alpha = R/(R + W)$$

where R is the number of correct (right) inferences made by the recognition heuristic computed across all pairs where one object is recognized and the other is not, and W is the number of incorrect (wrong) inferences under the same circumstances.

Could It Ever Be Smart to Reason by Recognition?

Food choice in rats may be guided by recognition, but what about inferences made by *Homo sapiens*? Won't inferences based on recognition (or more fittingly, on ignorance) be little more than guesses? Consider two examples of people using the recognition heuristic.

Which U.S. City Has More Inhabitants: San Diego or San Antonio? We posed this question to students at the University of Chicago and the University of Munich. Sixty-two percent of the University of Chicago students, who have a reputation for being among the most knowledgeable in the United States, chose the correct answer. However, 100% of the German students chose correctly. How did the Germans infer that San Diego was larger? All of the German students had heard of San Diego, but many of them did not recognize San Antonio. They were thus able to apply the recognition heuristic and make a correct inference. The American students, recognizing both cities, were not *ignorant* enough to be able to apply the recognition heuristic.

Which English Soccer Team Will Win? Fifty Turkish students and 54 British students made forecasts for all 32 English F.A. Cup third-round soccer matches (Ayton & Önkal, 1997). The Turkish participants had very little knowledge about English soccer teams, while the British participants knew quite a bit. Nevertheless, the Turkish group made predictions that were nearly as accurate as those of the English group (63% versus 66% correct). English soccer teams are usually named after English cities (for example, Manchester United), and people who are ignorant of the quality of English soccer teams can still use city recognition as a cue for soccer team performance. Cities with successful soccer teams are likely to be large, and large cities are likely to be recognized. Empirical evidence indicates the Turkish students indeed used the recognition heuristic: Among the pairs where one team was recognized (familiar to some degree) but the other was not, the former team was chosen in 627 out of 662 cases (95%). As before, the recognition heuristic can turn partial ignorance into reasonable inferences.

Both studies illustrate the ecological rationality of the recognition heuristic. The recognition heuristic is ecologically rational in the sense that it exploits the structure of information in natural environments: Lack of recognition in these environments is systematic and not random. Ignorance is beneficial if it is correlated with what one wishes to infer. The heuristic is not a general-purpose strategy because this correlation holds in some situations, but not in all. In many environments involving competition, such as inferring which of two colleges is more highly ranked, or which of two teams will win a match, the recognition heuristic works well. However, there are tasks in which recognition is not a good predictor. Let us look more closely at when the recognition heuristic succeeds and fails.

Accuracy of the Recognition Heuristic

What is the proportion of correct answers one can expect to achieve using the recognition heuristic on two-alternative choice tasks? Suppose there is a reference class of N objects and a test consisting of pairs of randomly drawn objects. When drawing pairs of objects, there are three ways they can turn out: one recognized and one unrecognized, both unrecognized, or both recognized. Suppose there are n recognized objects and thus $N - n$ unrecognized objects. This means that there are $n(N-n)$ pairs where one object is recognized and the other is unrecognized. A similar calculation shows that there are $(N-n)(N-n-1)/2$ pairs in which neither object is recognized. Finally, there are $n(n-1)/2$ pairs where both objects are recognized. To transform each of these absolute numbers into a proportion of cases, it is necessary to divide each of them by the total number of possible pairs, $N(N-1)/2$.

To compute the proportion correct on such a test, it is necessary to know the probability of a correct answer for each type of pair. Recall that the recognition validity α is the probability of getting a correct answer when one object is recognized and the other is not. The probability of getting a correct answer when neither object is recognized (and a guess must be made) is .5. Finally, let β be the *knowledge validity*, the probability of getting a correct answer when both objects are recognized. Combining all these terms together, the expected proportion of correct inferences, $f(n)$, on an exhaustive pairing of objects is:

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

The right side of the equation breaks into three parts: the leftmost term equals the proportion of correct inferences made by the recognition heuristic; the middle term equals the proportion of correct inferences resulting from guessing; the rightmost term equals the proportion of correct inferences made when knowledge beyond mere recognition can be used. Inspecting this equation, we see that if the number of cities recognized, n, is 0, then all questions will lead to guesses and the proportion correct will be .5. If $n = N$, then the leftmost two terms become zero and the proportion correct will be β. We can also see that the recognition heuristic will come into play most when the participant is operating under "half ignorance," that is, when half of the objects are recognized ($n = N - n$), because this condition maximizes the number of pairs $n(N-n)$ in which one object is recognized and the other is unrecognized.

To summarize, based on the recognition validity α, the knowledge validity β, and the degree of ignorance, that is, n compared to N, Equation (1) specifies the proportion of correct inferences made by someone who uses the recognition heuristic. Now we will look at the most counterintuitive property of the recognition heuristic: the less-is-more effect.

The Less-Is-More Effect

Imagine that MacAlister's three sons have to take a quiz at school about German cities. The quiz consists of randomly drawn, two-alternative questions about population sizes of the 50 largest German cities. The youngest brother is ignorant, and has never even heard of Germany (not to speak of German cities) before. The middle brother is savvy, and recognizes 25 of the 50 largest cities from what he has overheard from day to day. The cities this middle brother recognizes are larger than the cities he does not recognize in 80% of all comparisons, that is, his recognition validity α is .8. The eldest brother is quite the scholar and has heard of all of the 50 largest cities in Germany. When any of the brothers recognizes both cities in a pair, he has a 60% chance of making the correct choice, that is, β is .6.

Suppose that all brothers use the recognition heuristic whenever they can. Which one will score the highest on the quiz? Figure 2-3, calculated from Equation (1), shows the performance of the three brothers. The smooth line connecting the points graphs the continuous version of Equation (1).

The youngest brother performs at chance level, and the eldest does better with 60% correct. Remarkably, the middle brother, who knows less than the eldest, makes the most accurate inferences. He is the only brother who can use the recognition heuristic. Moreover, he can make the best of his ignorance because he happens to recognize half of the cities, and this allows him to use the recognition heuristic most often. The recognition

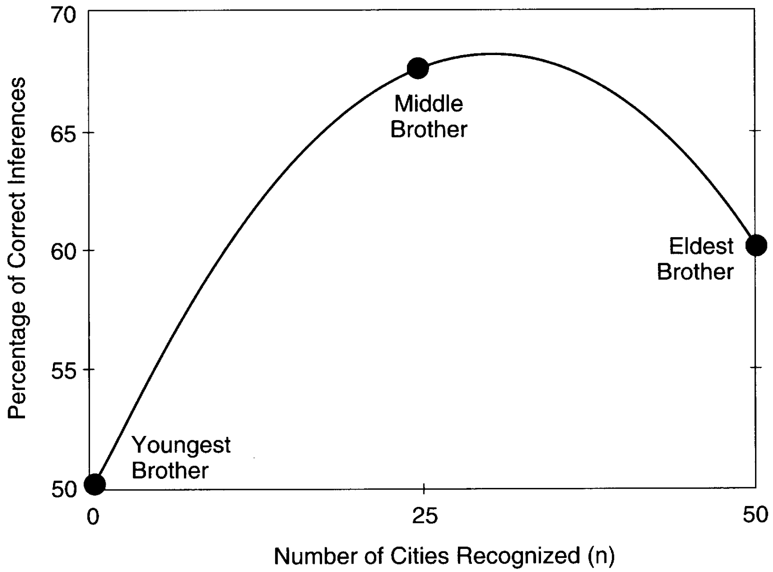

Figure 2-3: An illustration of a less-is-more effect. The youngest brother has never heard of any German city, and performs at chance level. The middle brother recognizes half of the 50 cities, and thus can apply the recognition heuristic in about half of the questions. This allows for 67.5% correct inferences (calculated from Equation (1); $\alpha = .8$ and $\beta = .6$). The oldest brother, who has heard of all the cities and thus knows more than the middle brother, gets only 60% correct inferences—a less-is-more effect. The curve also shows the performance for intermediate states of lack of recognition (calculated from Equation (1)). Note that the curve does not peak over the middle brother, but rather has its maximum slightly to the right of him. The reason for this is that β is .6 rather than .5.

heuristic can thus lead to a paradoxical situation where those who know more exhibit lower inferential accuracy than those who know less.

When Will the Less-Is-More Effect Occur?

The situation in which the less-is-more effect occurs can be stated in general terms. In the type of two-alternative tests described here where the recognition heuristic is consistently applied, *a less-is-more effect occurs when the recognition validity* α *is greater than the knowledge validity* β.

If this condition does not hold, then inferential accuracy will increase as more and more objects become recognized. We derive this result mathematically in Goldstein and Gigerenzer (1998).

A mathematical demonstration, however, is always based on simplifying assumptions. Here, for example, we have supposed that the recognition validity α remains constant across the x-axis in figure 2-3. In contrast to this figure, which represents individuals (the brothers) with different knowledge states and fixed α, the recognition validity usually varies when one individual comes to recognize more and more objects. The intuition for this result is as follows. When there are many different individuals with various levels of recognition, it is possible that each individual has the same recognition validity (that is, the objects they recognize are larger than the objects they do not recognize a certain proportion of the time, which we call α). However, when one individual comes to recognize more and more objects, the recognition validity changes because each newly recognized object, depending on how large it is, will increase or decrease the recognition validity. That is, coming to recognize smaller objects decreases recognition validity, and coming to recognize larger objects increases it.

Thus, the question must be posed: Can we demonstrate a less-is-more effect using realistic sequences of learning that do not satisfy the simplifying assumption that α is constant as n varies?

We created a computer program that learns about German cities in order of how well-known they are. To estimate this order, we surveyed 66 University of Chicago students and asked them to select the cities in Germany they recognized from a list, and then we ranked the cities by the number of people who recognized them. With this data, we hoped to approximate the order in which an American might learn about cities in Germany. The computer program first learned to recognize only Munich, the most well-known city, and was then given an exhaustive quiz consisting of all pairs of German cities. Next, already knowing Munich, it learned to recognize Berlin, the second most well-known city, and was tested again. It learned to recognize city after city until it recognized them all. In one condition, the program learned *only* the names of cities and made all inferences by the recognition heuristic alone. This result is shown by the bottom line on figure 2-4 labeled "no cues." When all objects were unrecognized, performance was at a chance level. Over the course of learning about cities, an inverse "U" shape appears, as in figure 2-3. Here the less-is-more curve is very jagged because, as mentioned, the recognition validity was not set to be a constant, but was allowed to vary freely as cities became recognized.

Would the less-is-more effect disappear if the computer program learned not just the *names* of cities, but information useful for predicting city populations as well? In a series of conditions with increasing information, the program learned the name of each city, along with one, two, or nine predictive cues for inferring population (the same cues as in Gigerenzer & Goldstein, 1996a). In the "one cue" condition, as the program learned to recognize a city, it also learned if it was once an exposition

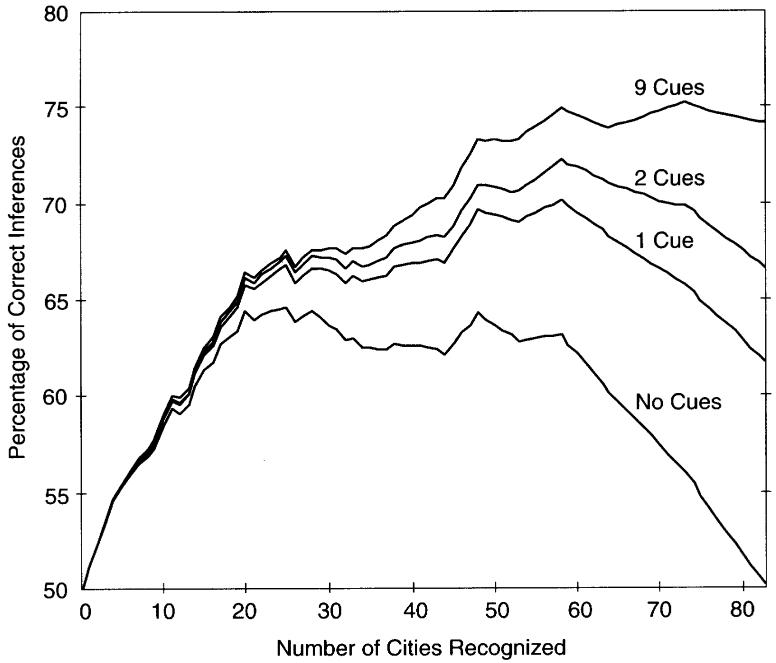

Figure 2-4: Less-is-more effects as cities become recognized in an order indicated by actual recognition data. Inferences are made on recognition alone (no cues), or with the aid of 1, 2, or 9 predictive cues.

site. Being an exposition site is a strong predictor of population with a high ecological validity of .91 (see Gigerenzer & Goldstein, 1996a).[1] The program then used a decision strategy called Take The Best (see chapters 4 and 5) to make inferences about which city is larger. All that is necessary to know for now is that Take The Best is an accurate strategy (as accurate as multiple regression for this task) for drawing inferences from cues, and it uses the recognition heuristic as its first step.

1. An ecological validity of .91 means that in 91% of the cases where one city has an exposition site and the other does not, the first is also the larger city. An ecological validity is a relation between a cue and a criterion, independent of a particular person. It is not the same as the knowledge validity β, which is the proportion of correct answers a person achieves when both objects are recognized, no matter what the values on the various cues are. Ecological validity is defined for the subset of pairs where both objects are recognized *and* one has an exposition site and the other does not. Both α and β are characteristics of a particular person.

Does adding predictive information about exposition sites wash out the less-is-more effect? It does not. The peak of the curve shifts slightly to the right, but it maintains its inverse "U" shape. When the program recognizes more than 58 cities, including information about exposition sites, the accuracy still goes down. In the "two cue" condition, the program learned if each city was an exposition site *and* if it had a soccer team in the major league—another cue with high validity (.87). The less-is-more effect was lessened—to be expected when adding knowledge—but still pronounced. Recognizing all cities and knowing all the information contained in two cues (the far right-hand point) resulted in fewer correct inferences than recognizing only 23 cities. Finally, in the "nine cues" condition, the program had all information about all nine cues available to it. This is surely more information for predicting German city populations than most German citizens know. This degree of knowledge must be enough finally to overcome the benefits of ignorance, right? Figure 2-4 shows the less-is-more effect finally flattening out. However, it does not go away completely: Even when all 747 (9 × 83) cue values are known and all cities are recognized, the point on the far right is still lower than more than a quarter of the points on that curve. A beneficial amount of ignorance can enable even higher accuracy than extensive knowledge.

The simulation can be summarized by two main results. The simplifying assumption that the recognition validity α remains constant is not a necessary precondition for the less-is-more effect. Moreover, the counterintuitive effect holds in this example even when complete knowledge about nine predictors is present.

A less-is-more effect can be observed in at least three different situations. First, it can occur between two groups of people, where the more knowledgeable group makes systematically fewer accurate inferences than a less knowledgeable group in a given domain. An example of this was the performance of the American and German students on the question about whether San Diego or San Antonio is larger. Second, a less-is-more effect can occur between domains, that is, where the same group of people makes a greater number of accurate inferences in a domain where they know little than in a domain where they know a lot. An empirical example will soon follow. Third, a less-is-more effect can occur over time, that is, where the same group makes increasingly worse inferences as they learn about a domain. For instance, the simulation results in figure 2-4 show how accuracy first increases and then decreases as knowledge is acquired.

So far, we have specified mathematically when the less-is-more effect occurs and shown that it also appears in realistic learning situations that violate the assumptions of the mathematical model. But can the effect be observed in real people? It could be that evolution has overlooked the inferential ease and accuracy the recognition heuristic affords. In the following section, we study whether people's judgments actually follow the recognition heuristic, and whether a less-is-more effect can be demonstrated empirically.

Empirical Evidence

Do People Use the Recognition Heuristic?

This simple test asks how often unprompted people will use the recognition heuristic. We quizzed Americans on all pairs of cities drawn from the 25 ($n = 6$) or 30 ($n = 16$) largest in Germany (300 or 435 questions) and asked them to choose the more populous city in each case. We had the participants check off from a list which of these cities they recognized, either before or after the test (this order, however, had no effect). From this recognition information, we could calculate how often participants had an opportunity to choose in accordance with the recognition heuristic and compare it to how often they actually did. Figure 2-5 shows the results for 22 individual participants. Note that the recognition heuristic predicts individual differences. Depending on the particular cities people recognize, their inferences about the population should systematically vary.

For each participant, two bars are shown. The darker bar shows how many opportunities the person had to apply the recognition heuristic, and the lighter bar shows how often that person's judgments agreed with the heuristic. For example, the person represented by the leftmost pair of bars had 156 opportunities to choose according to the recognition heuristic, and did so every time. The next person did so 216 out of 221 times and so on. The proportions of recognition heuristic adherence ranged between 100% and 73%. The median proportion of inferences following the recognition heuristic was 93% (mean 90%).

This simple test of the recognition heuristic showed that people adhere to it the vast majority of the time. Let us put the heuristic to a tougher test. Would people still rely on it when given information that suggests doing otherwise?

Do People Use the Recognition Heuristic Despite Conflicting Information?

In this experiment, we taught participants useful information that offered an alternative to following the recognition heuristic. The information was about the presence of major league soccer teams, powerful predictors of city population in Germany. We wanted to see which people would choose as larger: an unrecognized city, or a recognized city that they just learned has *no* soccer team. To get an idea of which German cities our participants might recognize, we ran a pilot survey of 26 participants and had them check off from a list those cities they had heard of before.

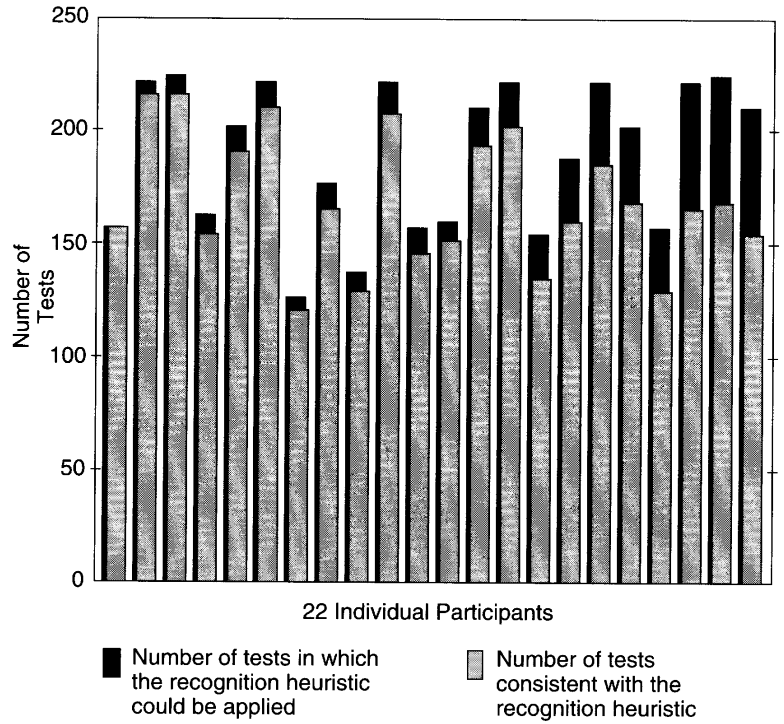

Figure 2-5: Recognition heuristic opportunities and usage by 22 individual participants. The individuals are ordered from left to right according to how closely their judgment agrees with the recognition heuristic. The darker bars are of different heights because individual participants recognized different numbers of cities.

The experiment began with a training session during which participants were instructed to write down all the information that would follow. They were first told that they would be quizzed on the populations of the 30 largest cities in Germany. Next they were taught that 9 of the 30 largest cities have soccer teams, and that the 9 cities with teams are larger than the 21 cities without teams in 78% of all possible pairs. Next, participants were allowed to draw eight cities at random and learn whether each has a soccer team or not. This drawing was rigged so that each participant chose the same four well-known cities that have soccer teams and four well-known cities that do not. Participants were then tested to make sure they could reproduce all of this information exactly, and could not proceed with the experiment until they did so. Either before or after the main task, participants were shown a list of German cities and asked to mark those that they had heard of before coming to the experiment.

With their notes beside them, participants were then presented pairs of cities and asked to choose the larger city in each pair. To motivate them to take the task seriously, they were offered a chance of winning $15 if they scored a high percentage correct. To reiterate, the point of the experiment was to see which participants would choose as larger: a city they have never heard of before, or one that they recognized beforehand but just learned has no soccer team. From the information presented in the training session (which made no mention of recognition), one would now expect a larger proportion of participants than in the previous experiment to choose the unrecognized city. Why? An unrecognized city either does or does not have a soccer team. If it does (a 5 in 22 chance from the information presented), then there is a 78% probability that it is larger, based on the soccer cue alone. If it does not, then soccer team information is useless and a wild guess must be made. The unrecognized city should be favored because any chance of it having a soccer team suggests that it is probably larger. Figure 2-6 shows the results.

The graph reads the same as figure 2-5. The darker bars are of different heights because individual participants recognized different cities before the experiment, so the number of cases where the recognition heuristic applied varied. Twelve of 21 participants made choices in accordance with the recognition heuristic without exception, while most others deviated on only one or two items. All in all, participants followed the recognition rule in 273 of the 296 total critical pairs. The median proportion of inferences agreeing with the heuristic was 100% (mean 92%), despite conflicting knowledge. These numbers are as high as in the previous experiment. It appears that the additional information was not integrated into the inferences, consistent with the recognition heuristic.

Does the Less-Is-More Effect Occur in Human Reasoning?

We have documented that the recognition heuristic can describe how humans make inferences in certain tasks. This result provides empirical support to the theoretical prediction that the less-is-more effect will appear. But we have yet to see this effect in the reasoning of real people. We administered two quizzes to 52 University of Chicago students. One quiz was on the 22 largest cities in the United States, cities about which they knew a lifetime of facts useful for inferring population. The other was on the 22 largest cities in Germany, about which they knew little or nothing beyond mere recognition—and they did not even recognize about half of them (Goldstein & Gigerenzer, 1998). Each question consisted of two randomly drawn cities, and the task was to pick the larger. One would expect American students to score substantially better on their native cities than on the foreign ones because of their lifelong acquaintance with their country. We considered this a tough test of the less-is-more effect. The curious phenomenon of a less-is-more effect is harder to demonstrate with real people than on paper, because the theory and simulation work we pre-

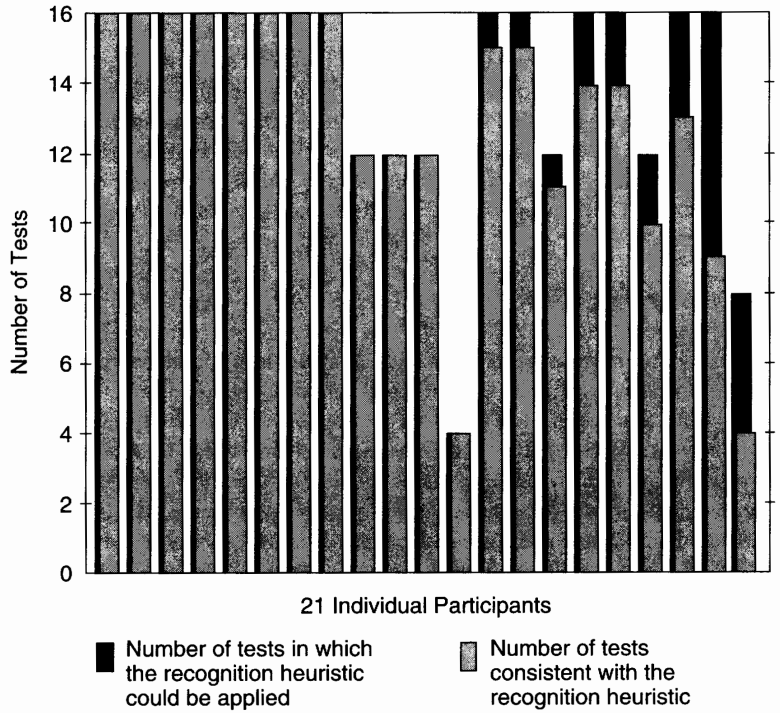

Figure 2-6: Recognition heuristic adherence despite training to encourage use of information other than recognition. The individuals are ordered from left to right according to how closely their judgment agrees with the recognition heuristic. The darker bars are of different heights because individual participants recognized different numbers of cities.

sented is about inference under uncertainty, but real people often have definite knowledge of the criterion. For instance, many Americans, and nearly all University of Chicago students, can name the three largest U.S. cities in order. This alone gives them the correct answer for 26% of the questions. Those who know the top five cities will get a free 41% correct. This definite knowledge of the rankings of the largest cities, combined with the lifetime of knowledge Americans have about their own cities, should make their scores on the domestic test hard to match.

The result was that the Americans scored a median 71% correct (mean 71.1%) on their own cities. On the less-familiar German cities, the median was a surprising 73% correct (mean 71.4%). Despite the presence of substantial knowledge about American cities, including some definite knowledge of which are the largest, the recognition heuristic resulted in a slight less-is-more effect. For half of the subjects, we kept track of which German cities they recognized, as in previous experiments. For this group, the median proportion of inferences according with the recognition heuristic was 91% (mean 89%). Furthermore, participants could apply the recognition heuristic nearly as often as possible, as they recognized a mean of 12 German cities, roughly half of the total. In a study that is somewhat the reverse of this one, a similar less-is-more effect was demonstrated with Austrian students who scored more accurate inferences on American cities than on German ones (Hoffrage, 1995; see also Gigerenzer, 1993).

Where Does Recognition Originate?

For some important adaptive tasks, such as avoiding food poisoning and identifying kin, organisms seem to be genetically prepared to act in accordance with the recognition heuristic. Wild Norway rats do not need to be taught to prefer recognized foods over novel ones. If a choice has life-threatening consequences, organisms that have to learn to use the recognition heuristic would likely die before they got the chance. Kin identification is an important adaptive task whose function seems to be avoiding incest and promoting nepotism (inclusive fitness) (Holmes & Sherman, 1983). Paper wasp females, for instance, use odor recognition (the odor they learned in their nest) to infer whether another wasp is a sister or nonsister. One can fool this mechanism by transferring newly emerged queens to a foreign nest, where they learn the odor of their (unrelated) nestmates (Pfennig et al., 1983). On the other hand, there are many domains in which organisms learn the predictive power of recognition through experience. Let us have a closer look at a source of name recognition in the realm of geography.

To what degree is the media responsible for the proper names we recognize? If the degree is high, the number of times a city is mentioned in the newspapers should correlate strongly with the proportion of the readers who recognize the city. The *Chicago Tribune* has a Sunday circulation of more than 1 million in the state of Illinois alone. We counted the number of articles published in the *Chicago Tribune* between 1985 and July 1997 in which the words "Berlin" and "Germany" were mentioned together. There were 3,484. We did the same for all cities in Germany with more than 100,000 inhabitants. The folks at the *Chicago Tribune* are not the world's most consistent spellers. We found Nuremberg spelled as Nurnberg, Nurnburg, Nuernberg, and Nuremburg (the database contained no umlauts). We searched under all the spellings we could imagine for all the cities. Table 2-1 illustrates that, for the top 12 German cities, the number of newspaper articles mentioning a city is a good predictor of whether its name will be recognized. What we call the surrogate correlation in figure 2-2, that is, the Spearman correlation (over all cities) between the number of newspaper articles mentioning a city and number of people recognizing it, is .79. But what about the actual populations? The ecological correlation, that is, the correlation between the number of newspaper articles and population, is .70. Finally, the correlation between the number of people recognizing the city's name and population is .60.[2]

Table 2-1: Recognition of German and American Cities

City	Articles	Recognition (%)	City	Articles	Recognition (%)
Berlin	3484	99	New York	493	100
Hamburg	1009	96	Los Angeles	300	100
Munich	1240	100	Chicago	175	97
Cologne	461	82	Houston	73	80
Frankfurt	1804	96	Philadelphia	67	63
Essen	93	28	San Diego	78	47
Dortmund	84	19	Phoenix	56	53
Stuttgart	632	63	Dallas	39	100
Düsseldorf	381	81	San Antonio	4	23
Bremen	140	44	Detroit	66	80
Duisburg	53	7	San Jose	13	17
Hannover	260	88	Indianapolis	20	50

Left side: Number of articles in 12 years of the *Chicago Tribune* mentioning the 12 largest German cities and the percentage of 67 University of Chicago students who recognized each city. Cities are ranked according to their actual size. *Right side:* Number of articles in 2 years of *Die Zeit* mentioning the 12 largest U.S. cities and the percentage of 30 University of Salzburg students who recognized each city.

These results suggest that individual recognition is more in tune with the media than with the actual environment, which indicates that city name recognition may come largely from the media. True population size is unknown to most people, but they can rely on mere recognition to make a fairly accurate guess.

But do these results stand up in a different culture? We looked at a major German-language newspaper, *Die Zeit*, and recorded the number of articles in which each of the U.S. cities with more than 100,000 inhabitants was mentioned. We compared this to the number of University of Salzburg students surveyed who recognized each city (Hoffrage, 1995). Table 2-1 shows again that the media references predict the number of people recognizing cities quite accurately. The surrogate correlation over all the cities between the number of newspaper articles and recognition is .86. The ecological correlation between the number of articles and population is .72, and that between recognition and the rank order of cities is .66. These results are quite consistent with those from the American participants, with slightly higher correlations. In all cases, the surrogate correlation is the

2. This correlation reflects the average recognition validity. It is calculated across persons, whereas the recognition validity is a characteristic of a particular person. The relation between validities and correlations is analyzed in chapter 6.

strongest, the ecological is the next strongest, and the correlation between recognition and the criterion is the weakest. In the next section, we see how institutions can exploit this relationship via advertising.

Institutions That Take Advantage of the Recognition Heuristic

Oliviero Toscani, the man behind the notorious Benetton advertising campaign, effectively bet his career on a series of advertisements that conveyed nothing about the product, but only sought to induce name recognition with shocking images such a corpse in a pool of blood, or a dying AIDS patient. In his book, Toscani (1997) reports that the campaign was a smashing success, which vaulted Benetton's name recognition higher than Chanel's and placed it among the top five brands in the world. Is recognition, regardless of how it is achieved, good for business? In the social world, name recognition is often correlated with wealth, resources, quality, power, and the like. Advertisers pay great sums for a place in the recognition memory of the general public. We have grown accustomed to seeing advertisements like Benetton's that communicate no product information besides proper names (this becomes especially clear visiting a foreign country where one has no idea to what the proper names refer). Less-known politicians, universities, cities, and even small nations go on crusades for name recognition. They all operate on the principle that if we recognize them, we will favor them.

There is evidence that one can induce name recognition furtively, and even unconsciously. The "overnight fame" experiments by Jacoby and colleagues (Jacoby, Kelley, Brown, & Jasechko, 1989; Jacoby, Woloshyn, & Kelley, 1989) demonstrate that people can be made confused about whether they have been shown a name in an experimental session or if they had encountered it before they came to the experiment. Jacoby's experiments have shown that exposing people to nonfamous names, waiting overnight, and then having them make fame judgments on these and other actually famous names causes them to confuse nonfamous names with famous ones. This demonstrates how a feeling of recognition can fool us into believing ordinary names are famous.

Mere Recognition Versus Degrees of Knowledge

We treat recognition as a binary phenomenon: one either recognizes or does not. How often one has been exposed to something is both hard to assess subjectively and irrelevant for the frugal recognition heuristic. These two features, the binary quality of recognition and the inconsequentiality of further knowledge, set the recognition heuristic apart from notions such as availability (Tversky & Kahneman, 1974), familiarity (Griggs & Cox, 1982), or the feeling of knowing (Koriat, 1993). The terms "availability" and "familiarity" are often used as common-sense explanations rather than as process models. Availability applies to items in memory and is often measured by the order or speed with which items come to mind, or the number of instances of a category that can be generated (see chapter 10). In contrast, as figure 2-1 shows, recognition concerns the difference between items in and out of memory (Goldstein, 1997). Availability is about recall, not about recognition. The term "familiarity" is typically used to denote a *degree* of knowledge or experience a person has with respect to a task or object. It does not pick up on the most important distinction for the recognition heuristic—that between recognized and unrecognized objects. As intuitive as notions such as availability and familiarity may be, there is a need to bring them from one-word explanations to precise models for heuristics (Gigerenzer, 1996). If this is done, then one could hope for a deeper, detailed understanding that can lead to unexpected consequences including the less-is-more effect.

A feeling of knowing, in Koriat's usage, is a person's assessment of the likelihood of being able to retrieve something from memory in the future. For example, the probe question "Who is the prime minister of Canada?" may put many non-Canadians into a tip-of-the-tongue state in which they may have a feeling about whether they will be able to retrieve the answer that is eluding them. Unlike the recognition heuristic, feelings of knowing presuppose knowledge beyond recognition, namely, the information held in the probe question. Another key difference is that the recognition heuristic can use recognition to predict some criterion in the world, whereas the feeling of knowing only predicts future memory performance.

The Recognition Heuristic as a Prototype of Fast and Frugal Heuristics

In this book, we study the architecture and performance of fast and frugal heuristics. The recognition heuristic is the simplest of these adaptive tools. It uses a capacity that evolution has shaped over millions of years, recognition, to allow organisms to benefit from their own ignorance. The heuristic works quickly and with limited knowledge—and even requires a certain amount of ignorance. The building blocks it uses for search, stopping, and decision are astoundingly simple. Search is limited to recognition memory—no recall of knowledge beyond recognition is attempted. Since search is limited in this way, the stopping rule is constrained—search terminates as soon as recognition has been assessed for both objects. The decision is consequently based on only one piece of information, recognition. Because a lack of recognition is essential for enabling a decision, we call this heuristic principle *ignorance-based decision making*. These heuristic principles add up to a conflict-avoiding strategy that eliminates the need for making trade-offs between cues pointing in different directions (as in the case where one recognizes a city but knows that it has no soccer team).

Fast and frugal heuristics, including the recognition heuristic, are based on psychological capacities such as recognition and heuristic principles such as ignorance-based decision making and one-reason decision making (relying on just one piece of information instead of aggregating several). The observation that people often try to avoid trade-offs and focus on one good reason has been documented numerous times (e.g., Baron, 1990; Hogarth, 1987; Payne et al., 1993). However, many scholars, psychologists included, have mistrusted the power of these heuristic principles, and saw in them single-mindedness and irrationality. This is not our view. The recognition heuristic is not only a reasonable cognitive adaptation because there are situations of limited knowledge in which there is little else one can do. It is also adaptive because there are situations, including those defined in this chapter, in which missing information results in more accurate inferences than a considerable amount of knowledge can achieve. In these situations, the recognition heuristic can be said to be ecologically rational, having the capacity to exploit structures of information in the environment in a simple and elegant way.