Satisficing Inference and the Perks of Ignorance

Daniel G. Goldstein and **Gerd Gigerenzer**
Center for Adaptive Behavior and Cognition
Max Planck Institute for Psychological Research
Leopoldstr. 24, 80802 Munich GERMANY
goldstein@mpipf-muenchen.mpg.de

Abstract

Most approaches to modeling rational inference do not take into account that in the real world, organisms make inferences under limited time and knowledge. In this tradition, the mind is treated as a calculating demon equipped with unlimited time, knowledge, and computational might. We propose a family of satisficing algorithms based on a simple psychological mechanism: one-reason decision making. These fast and frugal algorithms violate fundamental tenets of classical rationality, for example, they neither look up nor integrate all information. By computer simulation, we held a competition between the satisficing Take The Best algorithm and various more "optimal" decision procedures. The Take The Best algorithm matched or outperformed all competitors in inferential speed and accuracy. Most interesting was the finding that the best algorithms in the competition, those which used a form of one-reason decision making, exhibited a startling "less-is-more" effect: they performed better with missing knowledge than with complete knowledge. We discuss the less-is-more effect and present evidence of it in human reasoning. This counter-intuitive effect demonstrates that the mind can satisfice and seize upon regularities in the environment to the extent that it can exploit even the absence of knowledge as knowledge.

Toward Satisficing

How does an organism make inferences about unknown aspects of the environment? Three directions have been searched in the hope of an answer. The first, which we might call the computational demon approach, equates reasoning with extensive calculation. It applies to models of mind which describe basic cognitive processes, such as estimation, inference, or categorization, as resulting from sophisticated computations. Examples of this are models of estimation based on multiple regression, or models of foraging behavior based on Bayes' Theorem. How can the mind carry out such tough statistical problems that took millennia of cultural evolution to conquer? This is where the demon comes in. The computational demon, common to all such models, is a consultant to the reasoning agent, capable computing all possible futures based on its extensive and infallible memory of all things past. While this approach is flattering to the organism doing the reasoning, it may posit more computational power than is plausible to assume exists in ordinary minds. Yet, such models abound in human and animal psychology.

Another way to look at reasoning came about in the past few decades and has had a powerful impact on psychology ever since. This is the heuristics-and-biases approach (Kahneman, Slovic & Tversky, 1982), which suggests that reasoning is governed by simple heuristics that generally do the right thing, but that may be systematically and wholly misled. In principle, it is a good idea: do away with computational demons, and replace them with simple principles which may do the job equally well. A problem with the heuristics-and-biases approach comes in practice where most of the research focuses more on biases than heuristics and the heuristics offered are notoriously vague (Gigerenzer & Goldstein, in press).

The third and most promising view comes from Herbert Simon (1956). This view states that good reasoning can come about by simple algorithms that "satisfice". The word satisficing is a blend of the words satisfying and sufficing, and means just that: finding near-optimal solutions to difficult problems under the limited computational constraints of ordinary minds. As with the heuristics-and-biases approach, the computational demon is replaced with something more psychologically plausible, though here the resultant reasoning is quick and clean, as opposed to quick and dirty. Another good feature of Simon's satisficing idea is that it stems from a computational tradition which favors using algorithms as models, instead of just simple heuristics in isolation. Algorithms are easily coded up as computer programs that a researcher can use to put a model through its paces.

What are these simple, intelligent satisficing algorithms capable of making near-optimal inferences? How fast and how accurate are they? In this research, we look at the effectiveness of a satisficing algorithm that operates with simple psychological principles that satisfy the constraints of limited time, knowledge, and computational might. At the same time, it is designed to be fast and frugal without a significant loss of inferential accuracy since it can exploit the structure of environments. For instance, this algorithm uses the "recognition principle", a simple form of one-reason decision making, which seems at first a liability but turns out to an effective and efficient heuristic. In simulating this and other algorithms computationally, we came across a surprising "less-is-more" effect: a certain class of satisficing algorithms made better inferences under conditions of missing knowledge than with complete knowledge. This effect is discussed and its existence is proven to be tied to the recognition principle. We begin with the inference task we used to measure the effectiveness of various algorithms.

The Task

We deal with inferential tasks in which a choice must be made between two alternatives on a quantitative dimension. Consider the following example: Which city has a larger population? (a) Hamburg (b) Cologne. Assume that a subject does not know or cannot deduce the answer to the question, but needs to make an inductive inference from related real-world knowledge. How is this inference derived? How can we predict choice (Hamburg or Cologne) from a person's state of knowledge?

We assume that to make an inference about which of two objects has a higher value, knowledge about a reference class is searched. In our example, knowledge about the reference class "cities in Germany" could be searched. The knowledge could consist of probability cues. For instance, when making inferences about populations of German cities, the fact that a city has a professional soccer team in the major league ("Bundesliga") may come to a person's mind as a potential cue. That is, when considering pairs of German cities, if one city has a soccer team in the major league and the other does not, then the city with the team is likely, but not certain, to have the larger population. It may be useful to think of a knowledge state of a matrix of objects and cues.

	a	b	c	d
Recognition	+	+	+	-
Cue 1	+	-	?	?
Cue 2	?	+	-	?
Cue 3	-	+	?	?
Cue 4	?	-	-	?
Cue 5	?	?	-	?

Figure 1: Possible knowledge state for 4 objects (a-d), 5 cues, and recognition knowledge.

Figure 1 models a possible limited knowledge state of a person. Limited knowledge means that the matrix of objects by cues has missing entries (that is, objects, cues, or cue values may be unknown). She has heard of three German cities, *a*, *b*, and *c*, but not of *d* (represented by three positive and one negative "Recognition" values). She knows some facts (cue values) about these cities with respect to five binary cues. For a binary cue, there are two cue values, "positive" (e.g., the city has a soccer team), or "negative" (it does not). "Positive" refers to a cue value that signals a higher value on the target variable (for example, having a soccer team is correlated with high population). Unknown cue values are shown by a question mark. Since she has never heard of object *d*, the recognition value of *d* is negative and all its other cue values are necessarily unknown.

The Environment

We tested the performance of the Take The Best algorithm on how accurately it made inferences about a real-world environment (for a more complete description of these simulations, see Gigerenzer & Goldstein, in press). The environment was the set of all cities in Germany with more than 100,000 inhabitants (83 cities after German reunification), with population as the target variable. The model of the environment consisted of 9 binary ecological cues (such as the soccer team cue), and the actual 9 x 83 cue values.

Each cue had an associated *ecological validity* which is indicative of its predictive power. The ecological validity of a cue is the relative frequency, in a reference class, that objects with positive cue values have higher target variable values than objects with negative cue values (e. g., the relative frequency that cities with soccer teams are more populous than cities without teams in all possible pairs). The ecological validity of the 9 cues we chose ranged over the whole spectrum: from .51 (only slightly better than chance) to 1.0 (certainty).

We simulated subjects with varying degrees of knowledge about this environment. To model limited recognition knowledge, we created subjects who recognized between 0 and 83 German cities. For each of these types of subject, we created 500 simulated individuals, who differed randomly from one another in the particular cities they knew. The simulation needed to be realistic in the sense that people are more likely to recognize large cities than small ones. We performed a survey to get an empirical estimate of the actual covariation between recognition of cities and city populations. In a pilot study of 26 undergraduates at the University of Chicago, we found that the cities they recognized (within the 83 largest in Germany) were larger than the cities they did not recognize in about 80% of all possible comparisons. We refer to this value as the "recognition validity". This value was incorporated into our simulations by choosing sets of cities (for each knowledge state, that is, for each number of cities recognized) where the known cities were larger than the unknown cities in about 80% of all cases. Thus, the cities known by the simulated subjects had the same relationship between recognition and population as did those of the human subjects.

Algorithms

We held a competition in which five decision algorithms, specially designed for two-alternative inference tasks, were matched against each other in a contest. The winner would be the algorithm which made the most correct inferences in the least amount of time. The first competitor is called the *Take The Best* algorithm (Gigerenzer & Goldstein, in press), because its policy is "take the best, ignore the rest".

The Take The Best algorithm assumes a subjective rank order of cues according to their validities. We call the highest ranking cue the "best" cue. Here are the steps of the algorithm:

(1) Recognition principle: The recognition principle is invoked when the mere recognition of an object is a predictor of the target variable (here, population). The recognition principle states: if only one of the two objects is recognized, then choose the recognized object. If neither of the two objects is recognized, then choose randomly between them. If both of the objects are recognized, then proceed to Step 2.

(2) Search for the values of the best cue: For the two objects, retrieve the cue values of the best cue from memory.

(3) Discrimination rule: Decide whether the cue discriminates. The cue is said to discriminate between two objects if one has a positive cue value and the other does not.

(4) Cue substitution principle: If the cue discriminates, then stop searching for cue values. Else, go back to Step 2 and continue with the next best cue until a cue that discriminates is found.

(5) Maximizing rule for choice: Choose the object with the positive cue value. If no cue discriminates, then choose randomly.

One important feature of this algorithm is that search extends through only a portion of the total knowledge in memory (as shown by the shaded parts of Figure 1), and stops immediately when the first discriminating cue is found. Thus, the algorithm is well suited to situations of limited time or knowledge. A seemingly irrational feature of the algorithm is that it does not attempt to integrate information, but uses cue substitution instead. This idea of basing an entire decision on one single cue is what we call one-reason decision making. Note that the recognition principle (Step 1), is a form of one-reason decision making. We shall later see how this satisficing mechanism can actually improve inferential accuracy.

Testing the Algorithms

With the help of some of our colleagues and statistician friends, we created five, more traditional competitors to compare to the Take The Best algorithm. In *Tallying*, the number of positive cue values for each object is tallied across all cues and summed to give a score for each city. The city with the largest number of positive cue values is chosen. *Weighted Tallying* is identical to tallying except that the values added to each city's score are weighted by the respective ecological cue validities. The *Unit-Weight Linear Model* adds one point to a city's score for each positive value, but subtracts one point for each negative cue value, and chooses the city with the best score. Finally, the *Weighted Linear Model* is similar its unit-weighted counterpart, except that it adds and subtracts weighted values (the ecological cue validities) instead of whole points.

We tested how well subjects using the various algorithms did at answering questions of the kind, "Which city has more inhabitants? (a) Heidelberg (b) Bonn." Each of the 500 simulated subjects in each of the 84 types was tested on the exhaustive set of 3403 city pairs resulting in a total of 500 x 84 x 3403 tests, that is, about 143 million for each algorithm. The number of correct inferences, and the amount of cue values looked up were recorded for each subject and algorithm.

Results

The competition had two quite surprising results. First of all, even though the Take the Best algorithm used far less information than the other algorithms (on average, less than a third of all available cue values), it matched or outperformed all other algorithms in the proportion of correct inferences (Figure 2).

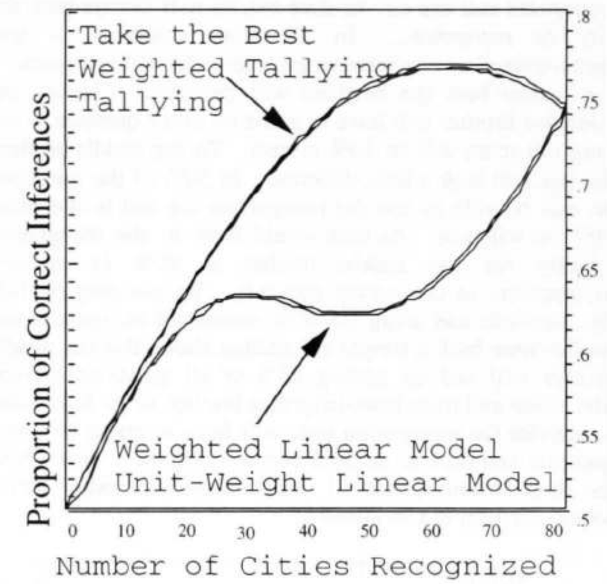

Figure 2: Less-is-more effect appearing for a variety of decision algorithms.

The second, very surprising result was that the best algorithms in the competition performed better with missing knowledge than with complete knowledge, as the non-monotonic upper curves suggest. Notice how for these curves, at any level of limited recognition knowledge of cities, learning more German cities will eventually cause a decrease in proportion correct. We call this intriguing finding the "less-is-more effect".

The Less-is-More Effect

What is behind the less-is-more effect? The most important factor is the recognition principle. All the algorithms which exhibit the less-is-more effect follow the recognition principle. In the case of Take The Best, it is a defining characteristic of the model, in the tallying variants, it arises as a side-effect. The linear models violate the recognition principle regularly; they often predict an unrecognized city to be larger than a recognized one (to understand why, see Gigerenzer & Goldstein, in press). Once this is realized, the reason for the effect can be seen analytically. We will build up to the more complicated analytic result from a simple thought experiment.

Imagine three brothers who sit down to take a quiz on the 100 largest German cities. The youngest brother is ignorant: he has never even heard of Germany before. The middle brother is savvy: he recognizes 25 of the 50 largest cities from what he has overheard from day to day. The eldest brother is quite the scholar: one day he took it upon himself to memorize the names of all the cities on his map of Germany. None of the brothers knows anything significant about the cities other than their names. Now suppose all three brothers adopt the same strategy for taking the test. Each decides that he will use the recognition cue wherever he can: in situations where he is given one city he recognizes and one city he does not, he will always pick the city he recognizes. In all other situations -- two unrecognized or two recognized cities -- he will just guess.

Consider how the brothers will perform. Clearly, the youngest brother will have to guess on every question -- his long-run score will be 50% correct. To the middle brother, the test will look a little different. In 50% of the questions he will be able to use the recognition cue and in the other 50% he will not. As luck would have it, the recognition validity for the middle brother is 80% (a realistic assumption, as our survey showed). By guessing on half the questions and using the 80% successful recognition cue on the other half, a simple calculation shows that the middle brother will end up getting 65% of all questions correct. The eldest and most knowledgeable brother, never being able to activate the recognition cue, will have to guess on every question and thus score 50% correct. Figure 3 shows how the three brothers, and all intermediate knowledge states, would perform in this domain.

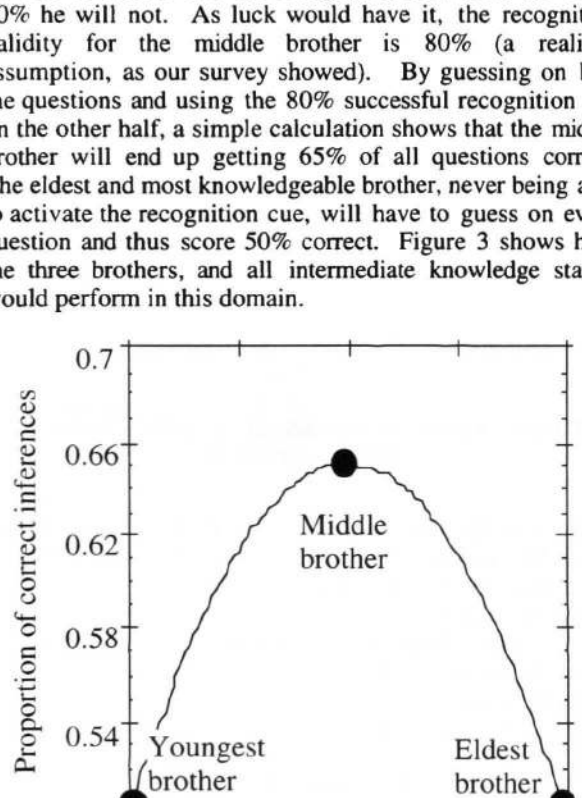

Figure 3: Performance for the three brothers and all intermediate knowledge states.

It becomes clear that the less-is-more effect is brought about by the variable applicability of the recognition principle in various knowledge states. When the recognition cue is not able to be activated, the system is forced to guess. This thought experiment may explain why an elder brother with no real knowledge beyond recognition may perform so poorly, but what about the simulation results, where a less-is-more effect persisted even when the simulated subjects had knowledge in the form of actual cue values? The following proof explains.

Suppose the number of objects (e.g., cities) in the reference class is N. Let n be the number of objects recognized by a subject. These two variables determine the proportions of question types that will appear on a quiz. In $2\left(\frac{n}{N}\right)\left(\frac{N-n}{N-1}\right)$ of all possible questions, one city will be recognized and the other not. In $\left(\frac{N-n}{N}\right)\left(\frac{N-n-1}{N-1}\right)$ of the questions, both cities will be unrecognized, whereas in $\left(\frac{n}{N}\right)\left(\frac{n-1}{N-1}\right)$ of the questions, both cities will be recognized.

When both cities are unrecognized, there is nothing to do but guess, and in the long run, half of these guesses will be correct. If one city is recognized and the other not, the recognition principle says to pick the recognized city. Let α be the probability of choosing the right answer via the recognition principle. If both cities are recognized, the inference has to be made using knowledge other than mere recognition. Let β be the probability of getting the right answer in this case. If α and β are both roughly constant and independent of n, the following function $f(n)$ gives the expected proportion of correct inferences:

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

By analyzing the graph of $\phi(n)$, the continuous version of $f(n)$, one sees that the less-is-more effect occurs if this curve has a maximum between $n = 0$ and $n = N - 1/2$. Solving the equation $\phi'(n) = 0$, when $\phi'(n)$ is simply the first derivative of $\phi(n)$, one locates the maximum of $\phi(n)$ at:

$$\frac{-(1 - 2\beta - 2N + 4\alpha N)}{2(1 - 4\alpha + 2\beta)} \qquad (*)$$

A simple calculation shows that when $\alpha=\beta$, the location of the curve's maximum is equal to N-1/2. Either increasing α or decreasing β from this point causes the fraction (*) to decrease, which implies the maximum of $\phi(n)$ will be displaced to the left. From this, we can conclude that there will be a less-is-more effect whenever $\alpha > \beta$, that is, whenever the accuracy of mere recognition is greater than the accuracy achievable when both objects are recognized.

Discussion

Two surprising results came out of this competition between algorithms. One is that a non-standard, satisficing algorithm performed as well as or better than all other algorithms in the competition, while looking up only one-third of the knowledge used by the competitors. The second was that the best algorithms in the competition did better with missing knowledge than with complete knowledge. The strong force most accountable for both these results was the simple and bold recognition principle, a form of one-reason decision making. The first result is an existence proof that a satisficing algorithm can do as well computationally-expensive models that use more information. The second suggests that, given the correct environment and an organism that follows the recognition principle, the less-is-more effect ought to emerge in the real world.

How difficult is it to find a less-is-more effect in real human behavior? Doing so depends on finding an environment with the correct structure, and people who apply the recognition principle. It is a relatively simple matter to find a reference class where a less-is-more effect should occur: simply find one where recognition is a better predictor than the environmental cues, and where the ecological validities of recognition and the cues do not change drastically as a more and more objects become recognized. This information about environment structures could be obtained from surveys and interviews about what objects people recognize, and how good they are at making inferences about these objects. Several experimental studies (Goldstein, 1994; Goldstein & Gigerenzer, 1996), show subjects exhibit a high degree of recognition principle adherence, even in cases where they are given information which suggests doing otherwise. The analytic results we have derived allow one to predict when and to what extent the less-is-more effect will occur.

The results of this study paint a new picture of the mind, not a picture where the mind is a computational demon, and not one where it is doomed to follow shoddy heuristics that lead it astray. Rather, it paints the mind as a time-pressed scavenger, one which uses the structures of natural environments to the degree that it can depend on a single, well-chosen cue as opposed to the costly aggregation of many, and one that can exploit any information -- even the very absence of knowledge -- to make accurate inferences about unknown features of the world.

Acknowledgments

Much of this research was completed at The University of Chicago in partial fulfillment of the first author's Master of Arts degree. We would like to thank Jean Czerlinski and Laura Martignon for their assistance on this paper.

References

Gigerenzer, G., & Goldstein, D. G. (in press). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*.

Goldstein, D. G. (1994). The less-is-more effect in inference. M. A. thesis, The University of Chicago.

Goldstein, D. G., & Gigerenzer, G. (1996). The recognition principle and the less-is-more effect. Manuscript submitted for publication.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129-138.