CHAPTER 11

More Is Not Always Better: The Benefits of Cognitive Limits

Ralph Hertwig
Max Planck Institute for Human Development, Berlin, Germany
and
Peter M. Todd
Max Planck Institute for Human Development, Berlin, Germany

THE LIABILITY VIEW OF COGNITIVE LIMITATIONS

Some of us ordinary mortals achieve extraordinary intellectual feats. For instance, the ancient Mithridates the Great (king of Pontus, a long and narrow strip of land on the southern coast of the Black Sea) is said to have learnt 22 languages, and to have been able in the days of his greatest power to transact business with the deputies of every tribe subject to his rule in their own peculiar dialect. Napoleon is known to have dictated 102 letters to successive teams of perspiring secretaries almost without pause, as he prepared the final details for the launching of his devastating campaign against Prussia (Chandler, 1997). One of the most celebrated physicists of our time was Richard Feynman, who won the 1965 Nobel Prize in physics for his many contributions to his field, especially for his work on quantum electrodynamics. Beyond being a brilliant thinker, on the bongos Feynman supposedly could play 10 beats with one hand against 11 with the other (Feynman, 1999; try it—you may decide that quantum electrodynamics is easier).

Despite numerous examples of people with prodigious abilities that we might otherwise have thought impossible, much of cognitive psychology rests on the premise that human information-processing capacity is rather severely bounded. In the words of Kahneman, Slovic and Tversky (1982), "cognitive psychology is concerned with internal processes, mental limitations, and the way in which the processes are shaped by the limitations" (p. xii). According to Cowan (2001), "one of the central contributions of cognitive psychology has been to explore limitations in the human capacity to store and process information" (p. 2). The list of documented limitations is long and includes the now classic thesis

that the capacity of short-term memory is restricted to a limited number of chunks of information—"the magical number seven, plus or minus two" (Miller, 1956). Similarly, the ability to pursue multiple intentional goals at any one time (for example, driving a car, planning one's day at work and, simultaneously, listening to the latest scoop on the stock market) is thought to be restricted by a limited budget of strategic processing capacity (e.g., Shiffrin, 1988; Barsalou, 1992, ch. 4).[1]

The premise that information-processing capacity is limited is usually accompanied by another ubiquitous assumption, namely, that these limitations pose a liability. They constrain our cognitive potential, this assumption holds, barring us from performing feats such as quickly computing the square roots of large numbers in our heads or reciting by heart all the entries of the Manhattan telephone book. Even more sinister, though, these cognitive limits are not only accused of hindering performance but are also suspected of being the culprit behind lapses of reasoning. In fact, the link between cognitive limitations and reasoning errors can be found in such disparate research programs as Piaget's theory of the cognitive development of children (e.g., Flavell, 1985), Johnson-Laird's mental model theory (1983; Johnson-Laird et al., 1999), and Kahneman and Tversky's heuristics-and-biases program (e.g., Kahneman, Slovic & Tversky, 1982).

Piaget, for instance, suggested that the still-immature mind of the preoperational child commits lapses of reasoning such as egocentrism (that is, the inability to take the perspective of another person) and animism (that is, ascribing lifelike qualities to inanimate objects). Only when cognitive development has reached its peak are children finally able to think in ways akin to those of scientists (for example, reasoning in accordance with the rules of logic and probability theory). The heuristics-and-biases program made a related point about the detriments of cognitive limits, but by challenging precisely the final stage of Piaget's developmental trajectory. In this research program's view, reasoning abilities that reflect the laws of probability and logic are *not* part of the intuitive repertoire of the adult human mind (e.g., Slovic, Fischhoff & Lichtenstein, 1976). Instead, due to their limited cognitive capacities, adults need to rely on quick shortcuts, heuristics, when they reason about unknown or uncertain aspects of real-world environments. But this use of heuristics leaves adult human reasoning prone to "severe and systematic errors" (Tversky & Kahneman, 1974, p. 1124), some of them akin to the lapses in reasoning that Piaget's preoperational children suffered from (such as, violation of class inclusion; see Hertwig, 2000).

What inspires the close link often made between bounds in cognitive capacity and lapses of reasoning, even irrationality? One speculative answer is that inferring this link naturally follows from a particular vision of rationality still embraced by many social scientists. This vision defines rational judgment and decision making in terms of *unbounded* rationality (see Gigerenzer & Todd, 1999). Unbounded rationality encompasses decision-making strategies that have little or no regard for humans' cognitive limitations and so are unfettered by concerns about decision speed or processing complexity. Theoretical frameworks such as subjective expected-utility maximization are often mathematically complex and computationally intractable; thus, they picture—implicitly or explicitly—the mind as if it were a supernatural being possessing unlimited powers of reason, boundless knowledge and endless time. Possibly, it is here that the link between limitations and irrationality suggests

---

[1] There are important exceptions such as parallel distributed memory models that disregard limited processing capacities by, for instance, assuming that search for a piece of information occurs simultaneously across multiple locations. In addition, there is evidence that the amount of information that can be held and processed in working memory can be greatly increased through practice (Kliegl et al., 1987; Ericsson & Kintsch, 1995), thus putting very narrow estimates of capacity limits somewhat into perspective.

itself to psychologists: being mere mortals, humans do not possess supernatural mental powers. Operating within the bounds of our cognitive limitations, we therefore must fall short of the norms defined by models of unbounded rationality.

The goal of this chapter is to challenge this obligatory link between cognitive limitations and human irrationality. While not doubting that limits can exact a price, we will question their exclusively negative status. Specifically, we put forth the thesis that limitations in processing capacity, as well as in other resources such as knowledge, can actually *enable* rather than *disable* important adaptive functions (*Thesis 1*). Secondly, we demonstrate that decision-making strategies that take limitations into account need not be less accurate than strategies with little or no regard for those limitations (*Thesis 2*). That is, we will show that accurate decision making does not necessitate supernatural mental powers, and thus that cognitive limitations need not be equated with inferior performance. Finally, we will challenge the assumption that simple decision-making strategies have evolved in response to the cognitive limitations of the human mind. We suggest the reverse causality and submit the thesis that capacity constraints may in fact be a byproduct of the evolution of simple strategies (*Thesis 3*).

THESIS 1: COGNITIVE LIMITATIONS CAN ENABLE IMPORTANT COGNITIVE FUNCTIONS

Because human beings are not omniscient, limitations in our knowledge are a ubiquitous fact—we differ only with regard to the domains in which we are more or less knowledgeable. In this sense, limited knowledge is an inevitable property of the *database* from which we derive inferences. Limitations in our knowledge, however, can be beneficial. We begin with an example of how limitations in knowledge can enable people to use a simple strategy to make surprisingly accurate inferences and predictions. But it is not just the data on which we base our decisions that are often limited—the *hardware* that we use to process those data and reach our conclusions is bounded as well. Hardware limitations—for instance, in terms of a limited working memory—also need not be a liability. In fact, as the later examples show, the limited capacity of human working memory can actually benefit learning and the vital inferences we make.

The Benefit of Limited Knowledge: The Recognition Heuristic

Most parents want their children to attend a good college. Unfortunately, the overwhelming variety of institutions of higher education makes the prospect of comparing them a daunting one. Just think of the many hundreds of US liberal arts colleges. How does one find out which are the good ones, or even just decide which of two colleges is the better one? Surprisingly, (partial) ignorance about the options in question can actually help people to make good choices. To see how limits in knowledge—in this case about colleges—can actually be beneficial, imagine the following scenario. Nearing the end of high school, three friends deliberate their choices of colleges. Because they are good students, they have applied only to liberal arts colleges that are ranked among the top 50 in the country. Eventually, each of the friends ends up with the choice between two colleges: *A* must choose between Middlebury and Vassar, *B* between Oberlin and Macalester, and *C* between Barnard and

Lafayette. Faced with these difficult choices, the friends turn to their parents for advice. Here is what they are told.

Student *A*'s parents have just moved to the USA. Thus, they know next to nothing about American colleges. In fact, they do not even recognize any of the colleges' names, and thus they can only guess which of the alternatives may be the better one. *B*'s parents also come from abroad—but they have already had the chance to absorb some knowledge about the American college system. Specifically, they recognize the names *Middlebury, Oberlin* and *Barnard* but do not recognize the names of the other contenders. Having no other knowledge to go on, they tell the three friends to go with those recognized alternatives. Finally, the friends turn to *C*'s mother, who happens to be a college professor. She has a lot of detailed knowledge about the colleges in question and can provide a plethora of information including the colleges' academic reputation, financial resources, student-to-faculty ratio, graduation rate and so on. Pressed to answer the question of which is the better college in each pair of choices, she responds: "It all depends!"

Although we all can resonate with the ideal that *C*'s mother represents, that knowing more about the alternatives in question is always better, such knowledge of multiple dimensions can create a predicament. In contrast to the convenient "common currency" assumption made by standard models of optimization (for example, translating everything into some amount of subjective expected utility), sometimes there is no way to compare all desires. Some things are incommensurable and thus difficult or impossible to convert into a single currency (Elster, 1979). For instance, should student *B* go to Oberlin because it has the higher academic reputation, or to Macalester because freshmen are more likely to return to campus the following year and eventually graduate (according to a recent college ranking published by *US NEWS*[2])? That is, should *B* strive to maximize the chance to get a good job or gain admission to a top graduate program, or should *B* try to maximize the chance of graduating by attending the school that may be offering the classes and services students need to succeed?

How can one escape this predicament of multiple, possibly incommensurable decision dimensions? Of course, one way to avoid it (later we will turn to another) is just to be ignorant about the intricacies of the choice situation—as was the case for *B*'s parents. But is this really a sensible path to take? Won't inferences based on pure recognition (and thus ignorance about other dimensions) be little more than random guesses? In fact, they can be a lot more. According to Goldstein and Gigerenzer (1999, 2002), choices based on recognition alone can be surprisingly accurate if exposure to different possibilities is positively correlated with their ranking along the decision criterion being used. They suggested that this is the case in environments involving competition (such as among colleges, baseball teams or companies). The decision task they focused on is a simple and common one: choose from two options (for example, colleges) the one that has a higher value on some criterion (for example, which one is better older and more expensive). Akin to the strategy that *B*'s parents used, Goldstein and Gigerenzer proposed the recognition heuristic for this kind of task. Simply stated, this heuristic says: *If one of two objects is recognized and the other is not, infer that the recognized object has the higher value*.

This minimal strategy may not sound like much for a decision maker to go on, but there is often information implicit in the *failure* to recognize something, and this failure can be exploited by the heuristic. To find out how good the recognition heuristic would fare in our

---

[2] The college rankings can be found at *http://www.usnews.com/usnews/edu/college/rankings/natlibs/natliba2.htm*.

college example, we conducted a small-scale study in which we asked a group of Americans and a group of Germans (all familiar with the academic system of their own country) to indicate which of the 50 highest-ranked American liberal arts colleges (listed in the *US NEWS* reference ranking) they recognized. We expected to observe two effects: first, that the American group would recognize many more college names than the German group; second, that the recognition validity (that is, the percentage of correct choices among those pairs where one college is recognized and the other is not) would nonetheless be higher in the German group.

This is exactly what we found. With years of experience of the college system, the Americans recognized about three-quarters (75 percent) of the college names, while the Germans recognized slightly more than one-fifth (22 percent). In addition, we found that the average recognition validity was higher for the German group: .74 compared to .62 for the Americans. What this means is that if we had asked our participants to choose higher-ranking colleges out of pairs of college names, the Germans could have used the recognition heuristic to pick those they recognized over those they did not, and this would have resulted in reasonably good choices (58 percent correct). In contrast, the Americans, who recognized most college names, would have made fewer good choices (54 percent correct).

This sounds promising in theory, but do people actually use the recognition heuristic? Goldstein and Gigerenzer (1999, 2002) conducted a series of experimental studies that strongly suggested that the recognition heuristic is used. Consider an example. Which city has more inhabitants: San Diego or San Antonio? When students at the University of Chicago were asked to answer questions like this by picking the larger of two American cities (comparisons constructed from the 22 largest in the USA), they scored a median 71 percent correct inferences. Surprisingly, however, when quizzed on city pairs from the 22 largest cities in Germany, the same students increased their score to a median 73 percent correct inferences. This result is counterintuitive when viewed from the premise that more knowledge is always better. The students knew a lifetime of facts about US cities that could be useful for inferring population, but they knew little or nothing about the German cities beyond merely recognizing about half of them. The latter fact, however, is just what allowed them to employ the recognition heuristic to pick German cities that they recognized as larger than those they did not. The students could not use this heuristic for choosing between US cities, though, because they recognized all of them and thus had to rely on additional retrievable information instead. Goldstein and Gigerenzer referred to this surprising phenomenon as the "less-is-more effect" and showed analytically and empirically that an intermediate amount of (recognition) knowledge about a set of objects can yield the highest proportion of correct answers—knowing (that is, recognizing) more than this will actually *decrease* the decision-making performance. We will return below to knowledge beyond recognition and demonstrate that variants of the less-is-more effect also exist for other kinds of knowledge.

Common wisdom has it that more knowledge or information is always better and that ignorance stands in the way of good decision making. The recognition heuristic is a counterexample to this wisdom. It feeds on partial and non-random ignorance to make reasonable choices, and it works because our lack of recognition knowledge about, for instance, colleges, sports teams (Ayton & Önkal, 1997) and companies traded on a stock market (Borges et al., 1999), is often not random, but systematic and exploitable. Thus, it is limited knowledge that enables the success of this powerful and very simple decision heuristic.

The Benefit of a Limited Working Memory: Covariation Detection

Writers and scientists alike agree that "the impulse to search into causes is inherent in man's very nature" (Tolstoy, 1982/1869, p. 1168), and that "humans exhibit an almost obsessive urge to mold empirical phenomena conceptually into cause-effect relationships" (Pearl, 1988, p. 383). Whatever the reasons for this human "obsession" with causality, the key point for our discussion is that limitations in human cognitive capacity may lay the groundwork for inferences of causality in terms of the early detection of covariation. In a series of papers, Kareev (1995a,b; 2000; Kareev, Lieberman & Lev, 1997) advanced the argument that limitations of working-memory capacity force people to rely on small samples of information drawn from real-world environments (and from their long-term memory). Small samples of information, however, have a specific advantage: they maximize the chances for early detection of a correlation.

Kareev's argument runs as follows. To determine whether two variables covary (for example, does flight behavior trigger a predator's chase behavior), one typically relies on data sampled from one's environment (and prior expectations; see Alloy & Tabachnik, 1984). If the assessment of a covariation has to be made "on the fly", the limited capacity of working memory imposes an upper bound on the size of the information sample that can be considered at one time. What is the size of the working memory and consequently the size of the information sample from which inferences are drawn? As we all know, the classic estimate of short-term memory is 7 ± 2 chunks (Kareev uses the term "working memory", akin to the earlier concept "short-term memory", but see Baddeley, 2000, on the different meanings of the term "working memory"). Taking Miller's (1956) estimate as a starting point, Kareev et al. (1997; Kareev, 2000) suggested that the limited capacity of working memory increases the chances for early detection of a correlation.[3] Here is the rationale.

Drawing small data samples increases the likelihood of encountering a sample that indicates a stronger correlation than that of the population. To see why, imagine drawing many small data samples of two continuous variables (for binary variables, the argument works slightly differently; see Kareev, 2000). If, for each sample, one calculates the relationships between the two variables (the Pearson product-memory correlation) and plots the distribution of the correlation coefficients found in the samples, the resulting sampling distribution will have a characteristic shape. Unless the correlation in the population is zero, the sampling distribution of the correlation will be skewed, with both the *median* and the *mode* of the distribution more extreme than the population value (see Hays, 1963, p. 530). Moreover, the amount of skewedness is a function of the sample size: the smaller the sample, the more skewed the resulting distribution.

In other words, for small sample sizes, many more samples will exhibit a sample correlation higher than the correlation in the population. Thus, when drawing a random sample from a population in which a correlation exists, any random sample is more likely than

---

[3] In a recent review article, Cowan (2001) concluded that over 40 years after Miller's seminal paper, we are still uncertain about the nature of the storage limit. For instance, according to some theories, there is no limit in storage per se, but a limit on the time an item can remain in short-term memory without being rehearsed. Cowan also argued that the storage limit itself is also open to considerable differences of opinion, but concluded that "the evidence provides broad support for what can be interpreted as a *capacity limit of substantially fewer than Miller's 7 ± 2 chunks; about 4 chunks on the average*" (emphasis is Cowan's, p. 3).

not to indicate a correlation more extreme than that found in the population.[4] Thus, the limited working memory functions as an amplifier of correlations. Consistent with this thesis, Kareev et al. (1997) found that people with smaller working-memory capacity detected correlations faster and used them to make correct predictions better than people with larger working-memory capacity. Moreover, they observed that the detection of correlation improved when it was based on smaller samples.

This theoretical account and empirical observations suggest a new and interesting view of cognitive limitations in general. In Kareev's view, cognitive limitations in working memory are not a liability but, in fact, enable important adaptive functions such as the early detection of covariation. The ability to detect contingencies early seems particularly important in domains in which the benefits of discovering a contingency outweigh the costs of false alarms. (Note that the smaller the data sample from which the contingency is inferred, the greater the variability of the sampling distribution and, consequently, the danger of a false alarm.) Such domains include, for instance, threats in which misses would be extremely costly.

Another Benefit from a Limited Memory Span: Language Learning

Another domain where limitations are beneficial, possibly even a prerequisite for maximal success, is language learning. According to Newport (1990), lesser ability to process and remember form-meaning mappings in young children allows them to learn more accurately those mappings that they do acquire and then to build further upon these as language learning proceeds. Late language learners, in contrast, may falter when attempting to learn all at once the full range of semantic mappings with their mature mental capacities.

This situation has been studied concretely by Elman (1993) in a neural network model of language acquisition. When he tried to get a large, recurrent neural network with an extensive memory to learn the grammatical relationships in a set of several thousand sentences of varying length and complexity, the network faltered. It was unable to pick up such concepts as noun–verb agreement in embedded clauses, something that requires sufficient memory to keep embedded and non-embedded clauses disentangled. Instead of taking the obvious step of adding more memory to the model to attempt to solve this problem, though, Elman counterintuitively *restricted* its memory, making the network forget everything after every three or four words. He hoped in this way to mimic the memory restrictions of young children first learning language. This restricted-memory network could not possibly make sense of the long, clause-filled sentences it was exposed to. Its limitations forced it to focus on the short, simple sentences in its environment, which it did learn correctly, mastering the small set of grammatical relationships inherent in this subset of its input. Elman then increased the network's effective memory by forcing it to forget everything after five or six words. It was now able to learn a greater proportion of the sentences it was exposed to, building on the grammatical relationships it had already acquired. Further gradual enhancements of the

---

[4] This skewed distribution is related to the fact that correlation coefficients are truncated, with their absolute values not exceeding 1 or −1. Assume the correlation coefficient in the population is .8. Sample correlations can deviate in two directions from the population parameter: they can be larger or smaller. A deviation above, however, can at most be .4, while a deviation below can go as far as −1.8. To offset the (few) very large deviations in the "wrong" direction, there must be many more (smaller) deviations in the "right" direction. From this, it follows that one is more likely to encounter a sample correlation that amplifies the population value than a sample correlation that attenuates it.

network's memory allowed it ultimately to learn the entire corpus of sentences that the full network alone—without the benefit of starting small—had been unable to fathom.

Elman sees the restrictions of the developing mind as enabling accurate early learning about a small portion of the environment, which then provides a scaffold to guide learning and hypothesizing about the rest of the environment in fruitful, adaptive directions. Cognitive "constraints" are no longer a negative limitation of our (or our children's) ability to behave adaptively in our environment. Rather,

> the early limitations on memory capacity assume a more positive character. One might have predicted that the more powerful the network, the greater its ability to learn a complex domain. However, this appears not always to be the case. If the domain is of sufficient complexity, and if there are abundant false solutions [for example, local error minima in a neural network's solution space], then the opportunities for failure are great. What is required is some way to artificially constrain the solution space to just that region which contains the true solution. The initial memory limitations fill this role; they act as a filter on the input, and focus learning on just that subset of facts which lay the foundation for future success. (Elman, 1993, pp. 84–85)

Thus, a smaller memory span should not be seen as a *constraint* on language learning, but rather as an *enabler* of learning, as Cosmides and Tooby (1987, p. 301) have put it.

Let us conclude this section with a cautionary note. We should be careful not to extend these arguments automatically to every problem environment that humans face—language, after all, has evolved culturally to be something that our fast and frugal developing minds can readily learn. But further explorations beyond Kareev's and Elman's work should reveal other domains where limited memory enables rather than constrains inference or learning.

THESIS 2: COGNITIVE LIMITATIONS AND SIMPLE PROCESSING NEED NOT BE EQUATED WITH INFERIOR PERFORMANCE

Scientific theorizing, visions of rationality and common wisdom alike appear to share a mutual belief: the more information that is used and the more it is processed, the better (or more rational) the choice, judgment or decision will be. This belief is not just an inconsequential idea that people might have. It affects, for instance, how we set up our information environments. According to Andrew Dillon (1996), for instance, "the belief that enabling access to, and manipulation of masses of information . . . is desirable and will somehow increase learning (however measured) is ever-present in discussions on educational hypertext" (p. 31). In his view, however, "to date, the claims have far exceeded the evidence and few hypertext systems have been shown to lead to greater comprehension or significantly better performance. . . . This concern with vast information sources over real human needs betrays the technocentric values of its proponents even while they talk in user-centred terms" (p. 32).

What is the evidence that more information and more complex processing are, a priori, better, or, vice versa, that less information and less processing, a priori, impair performance? The research program that has most strongly advocated the view that less processing, via the use of simple cognitive heuristics (relying on simple psychological principles such as associative strengths), can yield severe and systematic errors is the heuristics-and-biases program

(Kahneman, Slovic & Tversky, 1982). Specifically, this program attributes numerous departures from classical probability norms in inductive reasoning—"cognitive illusions", such as overconfidence, base-rate neglect and the conjunction fallacy—to the application of heuristics (Kahneman & Tversky, 1996). Some have argued that these departures "should be considered the rule rather than the exception" (Thaler, 1991, p. 4), while others have shown that a simple change in the way statistical information is represented—from single-event probabilities to frequencies— substantially reduces those departures (e.g., Gigerenzer, 1991; Gigerenzer & Hoffrage, 1995; Hertwig & Gigerenzer, 1999; but see Mellers, Hertwig & Kahneman, 2001).

Are violations of rational norms really the rule, and is simple processing to be equated with inferior performance? Taken at face value, the research in the tradition of the heuristics-and-biases program suggests a positive answer. However, Kahneman and Tversky (1982, p. 124) themselves acknowledged that "although errors of judgments are but a method by which some cognitive processes are studied, the method has become a significant part of the message". It appears that as a consequence of the exclusive focus on errors, the original assessment of heuristics as "highly economical and usually effective" (Tversky & Kahneman, 1974, p. 1131) has been largely ignored, and research in the tradition of the heuristics-and-biases program has been silent on questions such as when and why simple heuristics yield good performance. Exactly these kinds of questions, however, are being addressed in a new research program that explores the performance of simple decision heuristics. The research program on fast and frugal decision heuristics (Gigerenzer, Todd & ABC Research Group, 1999) challenges the equation of simple processing and inferior performance.

Simple Heuristics That Make Us Smart

Earlier, we introduced one fast and frugal decision rule studied within this program, the recognition heuristic. It exploits the knowledge of whether or not an option (such as a college name) has ever been encountered before. Often, however, more than just this type of information is accessible. In what follows, we describe two more fast and frugal heuristics that can be applied if more than just recognition knowledge is available. To illustrate how they work, let us return to our introductory example—deciding which of two colleges is better.

How would a rational agent make this decision? Two commandments that are often taken as characteristics of rational judgments are *complete search* and *compensation* (see Gigerenzer & Goldstein, 1999). The former prescribes, "thou shalt find all the information available", while the latter says, "thou shalt combine all pieces of information" (that is, not rely on just one piece). Thus, to decide which college is better, the decision maker ought to retrieve all the information available (either from internal or external memories), and then somehow combine the pieces of information into a single judgment (typically, this implies that the information will first be weighted according to its predictive value for the decision criterion).

More or less the exact opposite of this "rational" approach is to rely on just a single dimension to make the decision. Such a strategy simultaneously violates the commandments of complete search and compensation. Here is how it would work. Imagine that the goal is to select one object (such as a college) from two possibilities, according to some criterion

on which the two can be compared (such as ranking). Several decision dimensions (cues) could be used to assess each object on the criterion.[5] A one-reason heuristic that makes decisions on the basis of a single cue could then work as follows:

(1) Select a cue dimension and look for the corresponding cue values of each object.
(2) Compare the two objects on their values for that cue dimension.
(3) If they differ, stop and choose the object with the cue value indicating a greater value on the choice criterion.
(4) If the objects do not differ, return to the beginning of this loop (step 1) to look for another cue dimension.

Such a heuristic will often have to look up more than one cue before making a decision, but the simple stopping rule (in step 3) ensures that as few cues as possible will be sought, thus minimizing the information-searching time taken. Furthermore, ultimately only a single cue will be used to determine the choice, minimizing the amount of computation that must be done.

This four-step loop incorporates two of the three important building blocks of simple heuristics (as described in Gigerenzer & Todd, 1999): a stopping rule (step 3) and a decision rule (also step 3—deciding on the object to which the one cue points). To finish specifying a particular simple heuristic of this type, we must also determine just how cue dimensions are "looked for" in step 1. That is, we must pick a specific information search rule—the third building block. Two intuitive search rules are to search for cues in the order of their ecological validity (that is, their predictive power with regard to the decision criterion) or to select cues in a random order. In combination with the stopping and decision rules described above, the former search rule makes up the Take the Best heuristic, and the latter makes up the Minimalist heuristic (Gigerenzer & Goldstein, 1996).

Both heuristics disobey the commandments of complete search and compensation. Could such an unorthodox approach possibly work? To answer this question, Czerlinski, Gigerenzer and Goldstein (1999) used a set of 20 environments to test the heuristics' performance. The environments varied in number of objects and number of available cues, and ranged in content from high-school dropout rates to fish fertility. The decision accuracy of Take the Best and Minimalist were compared to those of two more traditional decision mechanisms that use all available information and combine it in more or less sophisticated ways: multiple regression, which weights and sums all cues in an optimal linear fashion, and Dawes's rule, which tallies the positive and negative cues and subtracts the latter from the former.

How did the two fast and frugal heuristics fare? They always came close to, and often matched, the performance of the traditional algorithms when all were tested on the data they were trained on—the overall average performance across all 20 data sets is shown in Table 11.1 (under "Fitting"). This surprising performance on the part of Take the Best and Minimalist was achieved even though they only looked through a third of the cues on average (and decided with only one of them), whereas multiple regression and Dawes's rule used them all (see Table 11.1, "Frugality"). The advantages of simplicity grew in the more important test of generalization performance, where the decision mechanisms were tested on a portion of each data set that they had not seen during training. Here, Take

---

[5] Cues can be either binary (is the college in the northeast of the USA?) or continuous (what is the student–faculty ratio?). For practical purposes, continuous variables can be dichotomized (for example, by a median split).

Table 11.1 Performance of different decision strategies across
20 data sets

Strategy	Frugality	Accuracy (% correct)	
		Fitting	Generalization
Minimalist	2.2	69	65
Take the Best	2.4	75	73
Dawes's rule	7.7	73	69
Multiple regression	7.7	77	68

Performance of two fast and frugal heuristics (Minimalist and Take the Best) and two
linear strategies (Dawes's rule and multiple regression) across 20 data sets. The mean
number of predictors available in the 20 data sets was 7.7. "Frugality" indicates the
mean number of cues actually used by each strategy. "Fitting accuracy" indicates the
percentage of correct answers achieved by the strategy when fitting data (test set =
training set). "Generalization accuracy" indicates the percentage of correct answers
achieved by the strategy when generalizing to new data (cross-validation, where test set ≠
training set) (data from Czerlinski, Goldstein & Gigerenzer, 1999).

the Best outperformed all three other algorithms by at least 4 percent (see Table 11.1,
"Generalization").

To conclude, making *good* decisions need not rely on the standard rational approach of
collecting all available information and combining it according to the relative importance
of each cue—simply betting on one good reason, even one selected at random, can provide
a competitive level of accuracy in a variety of environments. Of course, not all choices
in life are presented to us as convenient pairs of options. Do the results on the efficacy of
simple heuristics hold beyond the context of deliberated choices? The answer is yes. Limited
processing of limited information can also suffice to perform such taxing tasks as estimating
a precise criterion value (see Hertwig, Hoffrage & Martignon, 1999) and choosing the one
category, from several possible, that a given object falls into (Berretty, Todd & Martignon,
1999). In short, psychological plausibility and precision are not irreconcilable, and simple
processing need not be equated with inferior performance.

THESIS 3: COGNITIVE LIMITATIONS MAY BE A BYPRODUCT OF THE EVOLUTION OF SIMPLE STRATEGIES

Although there is little dispute that we humans often employ simple shortcuts or heuristics
to reach decisions, there is much debate about *how* we use them—at our peril or to our
advantage (e.g., Kahneman et al., 1982; Chase, Hertwig & Gigerenzer, 1998; Gigerenzer
et al., 1999). An issue that seems equally important but, to date, has received hardly any
attention is this: *why* is our mental machinery equipped with simple heuristics in the first
place? One likely reason why this question is hardly addressed is that there is an apparently
convincing straightforward answer: we rely on simple heuristics not because we choose
to but because we have only limited processing capacities at our disposal. They, in turn,
dictate the use of strategies that do not overtax our precious processing resources. Payne,
Bettman and Johnson (1993), for instance, put this traditional argument very clearly: "Our
basic thesis is that the use of various decision strategies [including simple heuristics] is an

adaptive response of a limited-capacity information processor to the demands of complex task environments" (p. 9).

Why is this argument not necessarily as plausible as it appears at first glance? The reason (see also Todd, 2001) is that given sufficient adaptive pressure to succeed in complex tasks, evolution could have built complex and sophisticated information-processing structures so that human cognitive machinery would not need to rely on simple, sometimes erroneous, heuristics. In other words, cognitive limitations could have been circumvented over the course of evolution—certainly at a price, such as the considerable costs involved in bearing a large-headed, long-dependent human baby, or the costs of high-energy expenditure for maintaining the metabolism of a large brain. That a human mind, in theory, could have evolved to be less subject to bounds in its memory and processing capacity is evidenced both by the prodigious processing that evolution provided for the seemingly more elementary processes such as perception or motor coordination, and by the extraordinary abilities of a few exceptional individuals (some of whom we listed in the introduction; see also Sacks, 1995).

If, for a moment, we do not take cognitive limitations as a given, but conceive of cognitive capacity as a free parameter that has been adjusted in the course of evolution, then a bold alternative answer arises to the question of why humans are equipped with cognitive limitations. In contrast to the traditional view, heuristics may not be dictated by cognitive limitations; rather, the evolution of simple heuristics may have required the evolution of no more than a certain limited amount of cognitive capacity, namely, the amount that was needed to execute them. This view reverses the traditional causal direction—from limitations that lead to heuristics to heuristics that require a certain, limited amount of capacity. This argument, however, can work only if simple heuristics had a selective advantage over more complex cognitive strategies (that would have required more processing power). What could those advantage(s) have been? Being fully aware that any answer to this question is speculative, we suggest two plausible candidate advantages—speed and robustness.

The Importance of Speed

One of the most pressing concerns facing a variety of organisms in a variety of dynamic environmental situations is simply the passage of time. This pressure arises primarily through competition between organisms in two main ways. First, time is short: organisms have occasional speed-based encounters where the slower individual can end up at a serious disadvantage; for instance, being slowly digested by the faster. Second, time is money, or at least energy: beyond predator–prey or combative situations, the faster an individual can make decisions and act on them to accrue resources or reproductive opportunities, the greater adaptive advantage it will have over slower competitors.

The speed argument, however, faces an important objection. Speed is only a precious resource if one assumes that search for information and processing of the retrieved information occurs serially. If, however, our mental hardware operates in a parallel fashion, even extensive search for information and sophisticated processing of it can occur rapidly. In other words, in a parallel machine, time is not a limiting factor. How could the parallel processing argument be countered? While this argument may be valid (to the extent that our mind is a parallel machine) for processes within the mind, it is not applicable to processes outside the mind—in particular, the process of search for information (for example,

the values of an object on various cue dimensions) in external sources. On an individual level, search for information in our environment occurs serially (ignoring the fact that our different senses can search in a parallel fashion). From on this reasoning, it is possible that many human decision heuristics were selected to achieve speed by seeking to use as little information from the environment as they could get away with.

All this is not to say that the entirety of human thought can be or should be characterized by simple heuristics—humans are uniquely able to set aside such mental shortcuts and engage in extensive cogitation, calculation and planning—but that we spend much of our time not taking the time to think deeply.

The Importance of Robustness

Learning means generalizing from the known to the unknown. This process of generalization has an element of gambling because the known information has both inherent structure and noise. Only the inherent structure, however, generalizes beyond the known information, and therefore this is what a learning model (for example, a decision strategy) should capture. Computationally powerful strategies (such as neural networks and multiple regression) aim to build a model of the known territory that is as perfect as possible, and thus to incorporate and account for as much of the known data as possible. Such a strategy is extremely successful if the known territory is large compared to the unknown, and if the known data include little noise. If the known territory, however, is small or includes much noise, trying to capture the known as precisely as possible turns out to be costly. Why? Because it means reliance not only on the inherent structure but also on the idiosyncrasies of the specific known information.

Take the US presidential election in 2000 as an example. Let us assume that the known data comprised only the election outcome in Florida, while the outcomes in the other states had to be predicted. As we all remember vividly, the outcome of the election in Florida was subject to many variables—some of them undoubtedly meaningful beyond Florida (for example, socioeconomic variables and the ethnic composition of Florida's constituency); others were relevant (if at all) only in the context of Florida's election turmoil (for example, poorly drafted "butterfly" ballots in one county and the secretary of state's interpretation of her "discretion"). Although across all 50 US states there is likely to be no true causal relationships between the election outcome and the variables idiosyncratic to Florida, the Florida sample of known data may (erroneously) indicate such relationships. Any inference model that tried to incorporate these idiosyncrasies to predict the election outcomes in the other states would be in danger of impairing its predictive power. In other words, it would "overfit" the known data.

How does the problem of "overfitting" relate to our thesis, namely, that simple heuristics may have had a selective advantage over more complex cognitive strategies? The argument we submit is that simple models are less prone to overfitting because they are parsimonious, using only a minimum number of parameters and thus reducing the likelihood of fitting noise (see Martignon & Hoffrage, 1999). Of course, there is a limit to simplicity, and there is "ignorant" simplicity (as in the case of that Minimalist heuristic, which randomly selects cues) and "smart" simplicity (as in the case of the Take the Best heuristic, which searches for one non-compensatory good reason, assuming that the structure of information is skewed in a non-compensatory way).
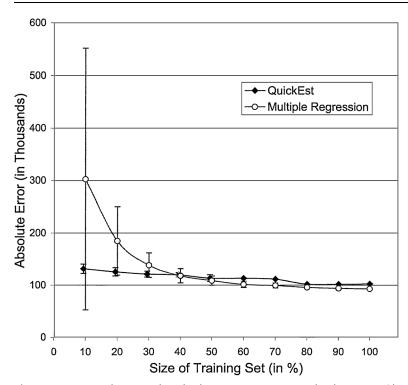
Figure 11.1 QuickEst's and multiple regression's mean absolute error (that is, absolute de-
viation between predicted and actual size) as a function of size of training set. Vertical lines
represent standard deviations. Note that some of the points have been offset slightly in the hor-
izontal dimension to make the error bars easier to distinguish, but they correspond to identical
training set sizes

Are fast and frugal heuristics, in fact, robust; that is, do they generalize well from known to
unknown territory? Using extensive simulations, we have consistently observed that simple
heuristics are, in fact, more robust than computationally complex strategies (e.g., Czerlinski
et al., 1999; Gigerenzer & Goldstein, 1999). Take the QuickEst heuristic (Hertwig, Hoffrage
& Martignon, 1999) as an example. The QuickEst heuristic is designed to estimate the values
of objects along some criterion (for example, how many people live in Maine?) using as
little information as possible. QuickEst does this by betting that the environment follows
a J-distribution, in which small values are common and big values are rare (here the "J"
is rotated clockwise by 90 degrees). Such distributions characterize a variety of naturally
occurring phenomena, including many formed by accretionary growth and phenomena
involving competition (such as scientific productivity).

How well would QuickEst do if it were to learn cues from a small sample? QuickEst
extracts from a learning sample only the order and sign of the cues, a very small amount of
information compared to the information extracted by complex statistical procedures such as
multiple regression (which extracts least-squares minimizing cue weights and covariances
between cues). Which is the better policy? Figure 11.1 shows QuickEst competing with
multiple regression at making generalizations from a training set to a test set. Each strategy
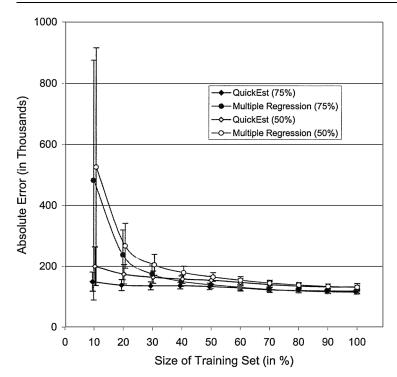estimated its respective parameters from a proportion (10 percent to 100 percent) of the

Figure 11.2 QuickEst's and multiple regression's mean absolute error (that is, absolute deviation between predicted and actual size) as a function of size of training set and of the amount of knowledge of cue values (75 percent and 50 percent). Vertical lines represent standard deviations

real-world environment of German cities with more than 100 000 inhabitants and values on eight ecological cues to population size (for example, is the city located in the industrial belt?) and made predictions about the complement.

Although (or because) QuickEst considers, on average, only 2.3 cues per estimate (out of 8 available cues), thus using only 32 percent of the information exploited by multiple regression, it exceeded the performance of multiple regression when the strategies had only scarce knowledge (that is, knew a third or fewer of the cities). When half of all cities were known, QuickEst and multiple regression performed about equally well. When the strategies had complete knowledge (all cities are known), multiple regression outperformed QuickEst by a relatively small margin. In other words, in the likely context of little to medium knowledge, QuickEst either matched the performance of multiple regression or outperformed it. Only when all knowledge was available—a situation that is rather unlikely to arise in most real-world domains—did multiple regression outperform QuickEst (by a small margin).

QuickEst's surprising performance is even more pronounced in a more difficult situation. Figure 11.2 shows the results for a simulation in which one-fourth or half of the cue values were eliminated from the environment (German cities) before the training and test sets were created, thus adding noise to the known data. Adding additional noise to the available information amplified QuickEst's edge over multiple regression. When only half of the cue

values were known, QuickEst outperformed multiple regression throughout the training sets (except for the 100 percent training set), and again the advantage was particularly pronounced when the training sets were small.

The performance figures for QuickEst and for other fast and frugal strategies (e.g., Czerlinski et al., 1999; Gigerenzer & Goldstein, 1999) demonstrate that, on these data sets, simple heuristics are less prone to overfitting a known environment and are thus more robust when generalizing to new environments than are more complicated statistical procedures such as multiple regression.

To conclude, in combination with their speed, robustness under conditions of limited knowledge may have provided simple strategies with a selective advantage over more complicated strategies. Cognitive limitations could thus be the manifestation of the evolutionary success of simple strategies rather than their origin.

CONCLUSION

In this chapter, we proposed a different view of the role of cognitive limitations. In this view, cognitive limitations (regarding knowledge and processing capacity) are not a nemesis—rather, they can enable important adaptive functions. Secondly, we demonstrated that decision-making strategies that take limitations into account need not be less accurate than strategies with little regard for those limitations. In opposition to the traditional view, according to which cognitive limitations dictate the use of simple heuristics, we finally proposed that some cognitive limitations may follow from the evolution of simple strategies.

There are different ways to think about and analyze the possible functions of cognitive limitations. One approach we did not pursue is to think about how a mind equipped with *boundless* capacities would function in the real world. Others, however, have taken this approach. Conducting a literary "*Gedanken Experiment*", the writer Jorge Luis Borges (1998) tells the story of Ireneo Funes, who, after a fall from a horse, found that his perception and memory had become essentially limitless. How did this man's perception of the world change as a function of his new abilities? Borges asserts that despite having an infinite memory, Funes is "not very good at thinking" (p. 137). Funes "was virtually incapable of general, platonic ideas . . . it irritated him that the 'dog' of three-fourteen in the afternoon, seen in profile, should be indicated by the same noun as the dog of three-fifteen, seen frontally" (p. 136). His mind consists of such perfect memory that no room exists for human creativity to link two dissimilar objects. In Borges' view, "to think is to ignore (or forget) differences, to generalize, to abstract" (p. 137), while Funes' memory is like a "garbage heap", (p. 135), which, whether he liked it or not, stored everything, the trivial and the important, indistinguishably.

Are these, in fact, the regrettable consequences of a perfect memory? Would we, as Borges suggests, become unable to function normally, because we were lost in a perpetual flux of unique instants of experience? Fortunately, we do not have to rely on our imagination (or on Borges' for that matter) to answer this question. There are real cases that approximate the *Gedanken Experiment* Borges engaged in; for instance, the wonderfully documented and fascinating case of S.V. Shereshevskii, a Russian memory-artist whose multisensory memory was studied over four decades by the neurologist A.L. Luria (1968/1987).

Resonating with Borges' portrayal of Funes, Luria described the most significant cost of possessing a memory that had "no distinct limits" (p. 11) as the inability to generalize, summarize and use abstractions. Shereshevskii told Luria: "I can only understand what I can visualise" (p. 130). He "was unable to grasp an idea unless he could actually see it, and so he tried to visualize the idea of 'nothing', to find an image with which to depict 'infinity'. And he persisted in these agonizing attempts all his life, forever coping with a basically adolescent conflict that made it impossible for him to cross that 'accursed' threshold to a higher level of thought" (p. 133).

To have "more memories than all mankind since the world began" (Ireneo Funes in Borges, 1998, p. 135) may not be so desirable after all.

ACKNOWLEDGEMENTS

REFERENCES

Alloy, L.B. & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, *91*, 112–149.

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

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Barsalou, L.W. (1992). *Cognitive Psychology: An Overview for Cognitive Scientists*. Hillsdale, NJ: Erlbaum.

Berretty, P.M., Todd, P.M. & Martignon, L. (1999). Categorization by elimination: Using few cues to choose. In G. Gigerenzer, P.M. Todd & the ABC Research Group (eds), *Simple Heuristics That Make Us Smart* (pp. 235–254). New York: Oxford University Press.

Borges, B., Goldstein, D.G., Ortmann, A. & Gigerenzer, G. (1999). Can ignorance beat the stock market? In G. Gigerenzer, P.M. Todd & ABC Research Group (eds), *Simple Heuristics That Make Us Smart* (pp. 59–72). New York: Oxford University Press.

Borges, J.L. (1998). *Collected Fictions*. New York: Penguin Books.

Chandler, D. (1997). Napoleon and death. *Journal of the International Napoleonic Society*, *1* (*http://www.napoleonseries.org/ins/scholarship97/c_death.html*).

Chase, V.M., Hertwig, R. & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, *2*, 206–214.

Cochran, B.P., McDonald, J.L., Parault, S.J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language*, *41*, 30–58.

Cosmides, L. & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupré (ed.), *The Latest on the Best: Essays on Evolution and Optimization* (pp. 277–306). Cambridge, MA: MIT Press/Bradford Books.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Science*, *24*, 87–185.

Czerlinski, J., Gigerenzer, G. & Goldstein, D.G. (1999). How good are simple heuristics? In G. Gigerenzer, P.M. Todd & ABC Research Group (eds), *Simple Heuristics That Make Us Smart* (pp. 97–118). New York: Oxford University Press.

Dillon, A. (1996). Myths, misconceptions and an alternative perspective on information usage and the electronic medium. In J.F. Rouet, J.J. Levonen, A.P. Dillon, & R.J. Spiro (eds), *Hypertext and cognition* (pp. 25–42). Hillsdale, NJ: Erlbaum.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71–99.

Elster, J. (1979). *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press.

Ericsson, K.A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Feynman, R.P. (1999). *The Pleasure of Finding Things Out: The Best Short Works of Richard P. Feynman*. Cambridge, MA: Perseus.

Flavell, J.H. (1985). *Cognitive Development*. Englewood Cliffs, NJ: Prentice-Hall.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology*, *2*, 83–115.

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

Gigerenzer, G. & Goldstein, D.G. (1999). Betting on one good reason: The Take the Best heuristic. In G. Gigerenzer, P.M. Todd & ABC Research Group (eds), *Simple Heuristics That Make Us Smart* (pp. 37–58*).* New York: Oxford University Press.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. Psychological Review, *102*, 684–704.

Gigerenzer, G. & Todd, P.M. (1999). Fast and frugal heuristics: The adaptive toolbox. In

Gigerenzer, G., Todd, P.M. & ABC Research Group (eds), *Simple Heuristics That Make Us Smart* (pp. 3–34). New York: Oxford University Press.

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

Goldstein, D.G. & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P.M. Todd & ABC Research Group (eds), *Simple Heuristics That Make Us Smart* (pp. 37–58). New York: Oxford University Press.

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

Hays, W.L. (1963). *Statistics for Psychologists*. New York: Holt, Rinehart & Winston.

Hertwig, R. (2000). The questionable utility of "cognitive ability" in explaining cognitive illusions. *Behavioral and Brain Science*, *23*, 678–679.

Hertwig, R. & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.

Hertwig, R., Hoffrage, U. & Martignon, L. (1999). Quick estimation: Letting the environment do the work. In G. Gigerenzer, P.M. Todd & ABC Research Group (eds), *Simple Heuristics That Make Us Smart* (pp. 209–234). New York: Oxford University Press.

Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M.S. & Caverni, J.-P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*, 62–88.

Kahneman, D. & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, *11*, 123–141.

Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, *103*, 582–591.

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

Kareev, Y. (1995a). Through a narrow window—working memory capacity and the detection of covariation. *Cognition*, *56*, 263–269.

Kareev, Y. (1995b). Positive bias in the perception of covariation. *Psychological Review*, *102*, 490–502.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review*, *107*, 397–402.

Kareev, Y., Lieberman, I. & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, *126*, 278–287.

Kliegl, R., Smith, J., Heckhausen, J. & Baltes, P.B. (1987). Mnemonic training for the acquisition of skilled digit memory. *Cognition and Instruction*, *4*, 203–223.

Luria, A.R. (1968/1987). *The Mind of a Mnemonist: A Little Book About a Vast Memory* (reprinted edn). Cambridge, MA: Harvard University Press.

Martignon, L. & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P.M. Todd & the ABC Research Group (eds), *Simple Heuristics That Make Us Smart* (pp. 119–140). New York: Oxford University Press.

Mellers, B, Hertwig, R. & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, *12*, 269–275.

Miller, G.A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11–28.

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

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann.

Sacks, O. (1995). *An Anthropologist on Mars*. New York: Vintage Books.

Shiffrin, R.M. (1988). Attention. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey & R.D. Luce (eds), *Stevens' Handbook of Experimental Psychology*. Vol. 2. *Learning and Cognition* (pp. 739–811). New York: Wiley.

Slovic, P., Fischhoff, B. & Lichtenstein, S. (1976). Cognitive processes and societal risk taking. In J.S. Carrol & J.W. Payne (eds), *Cognition and Social Behavior* (pp. 165–184). Hillsdale, NJ: Erlbaum.

Thaler, R. (1991). *Quasi Rational Economics*. New York: Russell Sage Foundation.

Todd, P.M. (2001). Fast and frugal heuristics for environmentally bounded minds. In G. Gigerenzer & R. Selten (eds), *Bounded Rationality: The Adaptive Toolbox* (*Dahlem Workshop Report*) (pp. 51–70). Cambridge, MA: MIT Press.

Tolstoy, L. (1982/1869). *War and Peace*. London: Penguin.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.