Rationality the Fast and Frugal Way

Nick Chater*, Martin Redington, Ramin Nakisa** & Mike Oaksford*****

*Department of Psychology
University of Warwick
Coventry, CV4 7AL, UK.
n.chater@warwick.ac.uk

**Dept. of Experimental Psychology
University of Oxford
South Parks Road, Oxford,
OX1 3UD, UK.
martin.redington@psy.ox.ac.uk
ramin@psy.ox.ac.uk

***School of Psychology,
University of Wales Cardiff, PO Box
901, Cardiff CF1 3YG, Wales, UK
oaksford@cardiff.ac.uk

Abstract

In a major theoretical paper, Gigerenzer and Goldstein (1996a) argue that classical rationality should be rejected as a norm of good reasoning, and that this thesis undermines both rational models of human thought and the alternative heuristics-and-biases program. They illustrate their argument by proposing that a specific cognitive estimation problem may be carried out by the "Take the Best" algorithm, which is "fast and frugal," but not rational. We argue: (1) that "fast and frugal" cognitive algorithms may *approximate* rational norms, and only in this way can their success be explained; and (2) that new computer simulations, and considerations of speed and generality, suggest that other algorithms are at least as psychologically plausible as Take the Best.

Introduction

Gigerenzer and Goldstein (1996; henceforth G&G; see also Goldstein & Gigerenzer, 1996) argue that human reasoning violates classical norms of rationality but nonetheless is adapted to the problems that it faces in the real world. Human reasoning is fast, frugal and effective—but not rational. They illustrate this proposal in the setting of a cognitive estimation problem: Deciding which is the larger of two cities, based on a list of features of each city. They present computer simulations comparing a very simple decision procedure, Take the Best, based on Gigerenzer's probabilistic mental models account (Gigerenzer, Hoffrage & Kleinbölting, 1991), with a range of alternative algorithms.

G&G have taken important steps in developing theories of "fast and frugal" reasoning, and provided a stimulating discussion of the relationship between human reasoning and rational norms. However, we believe their conclusions to be mistaken. In this paper, we put forward two challenges to G&G's arguments. First, we argue that human reasoning *must* be compared against rational norms, in order to explain why people's "fast and frugal" reasoning strategies are successful. Specifically, we suggest that G&G's radical rejection of classical rationality stems from conflating two levels of explanation: the level of *rational analysis* (Anderson, 1990, 1991a; Oaksford & Chater, 1994, 1995) where classical rationality holds; and the level of cognitive algorithms which are bounded by cognitive limitations, but which may serve as approximations to rational norms. Second, we present simulations comparing Take the Best against a range of alternative models which have been widely used in psychology or artificial intelligence. All these algorithms have similar levels of performance on the estimation problem G&G consider, which indicates that this problem does not usefully discriminate between cognitive algorithms. We argue that considerations of generality and speed suggest that other algorithms are at least as psychologically plausible as Take the Best.

Bounded Rationality and the Explanation of Human Inference

Almost all aspects of cognition involve *uncertain* inference, from word perception to learning to motor control. All of these inferences are provisional and uncertain, and may be revised in the light of more information (e.g., Oaksford & Chater, 1991). But, overall, human uncertain inference is spectacularly successful—the cognitive system vastly outperforms the most sophisticated artificial intelligence systems in almost every real-world domain. Explaining how this success is possible requires (1) specifying the cognitive algorithms underlying human uncertain inference. But it also requires (2) explaining *why* these algorithms lead to successful inference. This second issue is the center of controversy in G&G's paper. G&G argue that there are three possible viewpoints:

1. The "classical" view that the algorithms involved in human reasoning follows the laws of probability theory and statistics, which define normative canons for uncertain reasoning. This view is held to claim that the mind is "...a Laplacian demon equipped with unlimited time, knowledge, and computational might...carrying around the collected works of Kolmogoroff, Fisher, or Neyman..." (p. 650; all page references to G&G, 1996 unless otherwise stated)

2. The "heuristics and biases" program (e.g., Kahneman, Slovic & Tversky, 1982) which suggests that "human inference is systematically biased and error-prone, suggesting that the laws of inference are quick-and-dirty heuristics" (p. 650). G&G claim that this viewpoint "has retained the

normative kernel of the classical view" as defining the standard against which success in reasoning should be judged.

3. The "bounded rationality" approach, which assumes that the cognitive system must *satisfice* rather than optimize. Moreover, G&G state that this approach implies that "the minds of living systems should be understood relative to the environment in which they evolved *rather than* to the tenets of classical rationality..." (p. 651) (emphasis added). G&G claim that this viewpoint is a radical departure from both the classical view and its traditional opponents, because it rejects classical theories of rationality not only as descriptions of human reasoning, but also as normative standards. G&G illustrate and argue for this third position.

We believe that the bounded rationality approach should not be formulated as a radical *alternative* to the classical viewpoint or to the heuristics and biases programs. Rather it should be viewed as a *synthesis* of the insights of both approaches, and to be continuous with both.

What is Bounded Rationality?

We argue that bounded rationality, as its name implies, should be viewed as an approximation to (unbounded) rationality. That is, bounded rationality involves rationality, subject to *constraints*. These constraints typically involve resource limitations, including computational restrictions imposed by time and memory space restrictions, and more generally by the capacities of the computational architecture of the cognitive system.

Simon (1955) introduced the idea of bounded rationality to be contrasted with optimization-based models of individual behavior then being developed primarily within economics, operations research, and decision theory rather than in psychology. These models assumed that people can be viewed as choosing between courses of action to maximize their subjective expected utility, as scheduling multiple activities to maximize productivity and so on. Simon points out that these optimization problems are typically computationally intractable. Hence, assuming that thought is a kind of computation, people are not capable of such optimization; at best they can *satisfice*—find an acceptable, though typically not optimal, solution. Notice that the very idea of satisficing implies that there is some standard which is being approximated—maximum expected utility, productivity, and the like; but it also implies that to understand how well this standard is achieved, we must consider the mechanisms by which it is approximated.

Rationality may be bounded to various degrees, depending on the nature and severity of the constraints. At one extreme, a computational system exhibiting bounded rationality might carry out the same kind of calculation as would be required by an unbounded rational system, except that the calculations are simplified in some way. For example, in an optimization problem, the system might not attempt to find a global optimum, but might settle for the any solution which is satisfactory, according to some criterion; or probabilistic calculations might be simplified by making strong independence assumptions (e.g., Pearl, 1988). At the other extreme, a computational system might exhibit bounded rationality by relying on simple heuristics, which are easy to implement, and will typically (although by necessity not always) give an outcome in line with unbounded rationality. Such heuristics are familiar in cognitive modelling in a wide range of domains. For example, in models of problem solving, for example, heuristics range from general purpose "weak" heuristics, such as means-ends analysis (Newell & Simon, 1972) to specific heuristics tailored to the problem domain (in chess, for example, these might concern the values of the pieces or features of the board position conferring strategic advantage), and heuristics based on pattern-matching with large numbers of past exemplars (Chase & Simon, 1973). These heuristics are very easy to apply, and can provide useful approximations to the information that would be gained from the rationally justified, but computationally intractable, process of exhaustively searching the entire problem space.

At all points on this continuum of bounded rationality, understanding the system's performance requires specifying at least (i) what rational goal is being approximated; (ii) what algorithms are being used, and how they serve to approximate the rational goal (subject to the relevant constraints). These two aspects of the explanation of the system suggest how understanding a system as exhibiting *bounded* rationality can reconcile the intuitions behind the classical rationality and the heuristics and biases approaches. The specification of (i) takes the form of an (unbounded) rational analysis; and (ii) describes the particular algorithms, which may be the "fast and frugal" heuristics that Gigerenzer and Goldstein suggest, which approximate the rational goal. Thus bounded rationality can be seen as integrating the two previous approaches to understanding human uncertain reasoning, rather than as third option, opposed to both.

Notice that there *are* possible extreme positions, connected with the classical rationality and the heuristics and biases programs, which would not fit within the framework of bounded rationality. The first is that the cognitive system implements *un*bounded rationality - and hence part (ii) of the bounded rationality explanation is unnecessary. We know of no cognitive scientist who advocates this position— indeed, it would be untenable on purely computational grounds: The computational intractability of most rational accounts of thought, including probability theory, decision theory and logic (e.g., Paris, 1992; Reiner, 1995) implies that no physical device could implement unbounded calculations in the time-scales relevant for human cognition.

The second extreme position is that the cognitive system consists of algorithms, which do not approximate any rational standard—and hence part (i) of the bounded rationality explanation is unnecessary. It is not clear how many theorists would advocate this position, but we suspect that very few would do so. The very word "heuristic" implies a shortcut to achieving some goal by longer, optimal means. Theorists, such as Kahneman and Tversky, who describe cognitive processes as heuristics, are thereby implicitly recognizing that some kind of normative rational explanation is being approximated, if imperfectly (as G&G affirm in their characterization of the heuristics and biases program). Moreover, the very idea that human thought can be understood as *reasoning* rather than as a collection of uninterpreted *procedures* involves the assumption that some rational norms are being approximated (see Oaksford & Chater, 1995). Giving up the idea that thought involves reasoning has catastrophic implications, not just within psychology, but more broadly: assumptions of (approximate) human rationality are at the core of "rational choice" explanations in the social sciences (e.g., Elster, 1986), micro-economics, and appear to be underpin the attribution of meaning both to mental states and to natural language (Davidson, 1984).

G&G claim that the cognitive system is fast and frugal, but not rational. But we have seen that all or almost all theorists concerned with human or animal behavior, whether from the "classical rationality" or "heuristics and biases" viewpoints agree on the framework of *bounded* rationality, where this is understood as integrating rational and algorithmic styles of explanation. Thus they agree that the cognitive system is fast, frugal, *and*, contra G&G, rational. G&G provide no reason to deviate from this consensus.

Recognizing that developing rational accounts of cognition is an empirical enterprise, Anderson (1990, 1991) has proposed an important methodology for discovering a "rational analysis" of bounded rational systems. His methodology involves six steps:

1. Specify precisely the goals of the cognitive system.
2. Develop a formal model of the environment to which the system is adapted.
3. Make minimal assumptions about computational limitations.
4. Derive the optimal behavior function given 1-3 above.
5. Examine the empirical evidence to see whether the predictions of the behavior function are confirmed.
6. Repeat, iteratively refining the theory.

Notice that the first two steps explicitly address the two empirical factors determining the correct rational theory mentioned above: 1. deals with what the cognitive system takes the task to be, and 2. concerns the structure of the environment. The third point explicitly addresses the *bounded* character of cognitive processes. Because Anderson's emphasis is on the rational theory, rather than how the cognitive system approximates that theory, he emphasizes cases where these assumptions are "minimal," to place the burden of explanation as much as possible on the rational theory. Moreover, Anderson (in the tradition of economists and zoologists mentioned above) also recommends that theorists should push purely rational accounts as far as possible before introducing cognitive constraints, because these constraints are relatively poorly understood.

These steps are the basis on which the optimal behavior function is calculated, and compared against the empirical evidence. Note that Anderson is not committed to the idea of perfect optimality, rather than satisficing, with respect to this function, merely that it provides an ideal to which the cognitive system approximately conforms. Hence, Anderson's method of rational analysis is a blueprint for attempting to understand the "rational" theory underlying systems exhibiting bounded rationality, and has been fruitfully applied across a range of cognitive domains (e.g., Anderson, 1991). This approach to understanding bounded rational systems stands in direct contrast to the approach that G&G advocate, because it places a rational standard at the center of psychological explanation, rather than dispensing with such a standard entirely.

How Plausible is Take the Best?

We now turn from general questions concerning rationality, to the specific question of the plausibility of Take the Best as a psychological hypothesis concerning cognitive estimation. Specifically, G&G consider the problem of estimating answering questions such as "Which city has the larger population? (a) Hamburg (b) Cologne. They assume that the participant does not know the answer, but must decide on the basis of facts that they know about the two cities: e.g., whether the cities have a major professional soccer team, whether it is in East or West Germany). Because correlations can be found between some of these features and population size (e.g., major soccer teams tend to be in large cities), these features can be used as a basis for answering questions concerning which of a pair of cities is larger. Specifically, G&G's simulations involve answering these questions on the basis of nine binary features associated with each city.

G&G argue for the cognitive plausibility of their "fast and frugal" algorithm, Take the Best, the core of which is as follows. People are assumed to rank features in order of how reliable they determining city size. Comparison of two cities then involves running through each the features from the most to the least reliable, until a feature is found on which the two cities differ—this feature then determines which city is judged to be the larger. This algorithm thus makes a decision purely on the basis of a single feature, rather than integrating information from all features. This is one of the reasons that G&G claim that their algorithm is non-rational. G&G's original competition showed that this simple algorithm performed just as well in judging city populations as algorithms such as multiple linear regression, and algorithms based on tallying (see G&G, 1996 for details).

G&G argue that the good performance of Take the Best in their competition, and its speed, is evidence for its cognitive plausibility. In this section, we (1) present results of a new competition, between Take the Best and a range of general purpose learning algorithms from psychology and artificial intelligence, in which all algorithms obtain similar levels of performance; (2) we argue that these other approaches are to be preferred on grounds of their generality, in particular regarding their ability to integrate information; (3) we argue that G&G's claim that Take the Best is preferable to other algorithms on grounds of speed depends on implicit assumptions about the cognitive architecture. We conclude that general purpose learning algorithms are at least as plausible as Take the Best.

A New Competition

G&G's Take the Best algorithm is specially tailored for the city population problem. We compare its performance with three kinds of *general purpose* algorithms, which have all been used extensively in cognitive science and artificial intelligence research: (1) exemplar-based algorithms, which assume that people store previous examples, and judge new examples in relation to their similarity to stored examples; (2) multilayered, feedforward neural networks, trained by back-propagation; (3) the decision tree classifier.

One difference between G&G's simulations and those reported here is that we have assumed that people only have to compare between familiar cities. G&G also use a "recognition principle:" that cities which are recognized are assumed to be larger than cities which are not recognized. This principle could be combined with any of the algorithms in our competition and hence does not help distinguish between them. We have therefore left out this aspect of G&G's analysis below.

Representation of data. G&G represent each city as a vector of nine binary (0 or 1) cue values. To facilitate comparison between algorithms, we represented a pair of two cities by nine features representing the *difference* between the nine cue values for each city. For example, for the cities with features (1, 1, 0, 0, 1, 0, 0, 1, 1) and (1, 0, 1, 1, 1, 0, 1, 1, 0), the corresponding training pattern would be (0, 1, -1, -1, 0, 0, -1, 0, 1). A tenth value for each pattern indicated whether the population of the first of each pair was smaller, equal to, or larger than the population of the second. Taking all pairs of distinct cities in both orders yielded a possible 83 × 82 = 6806 training patterns.

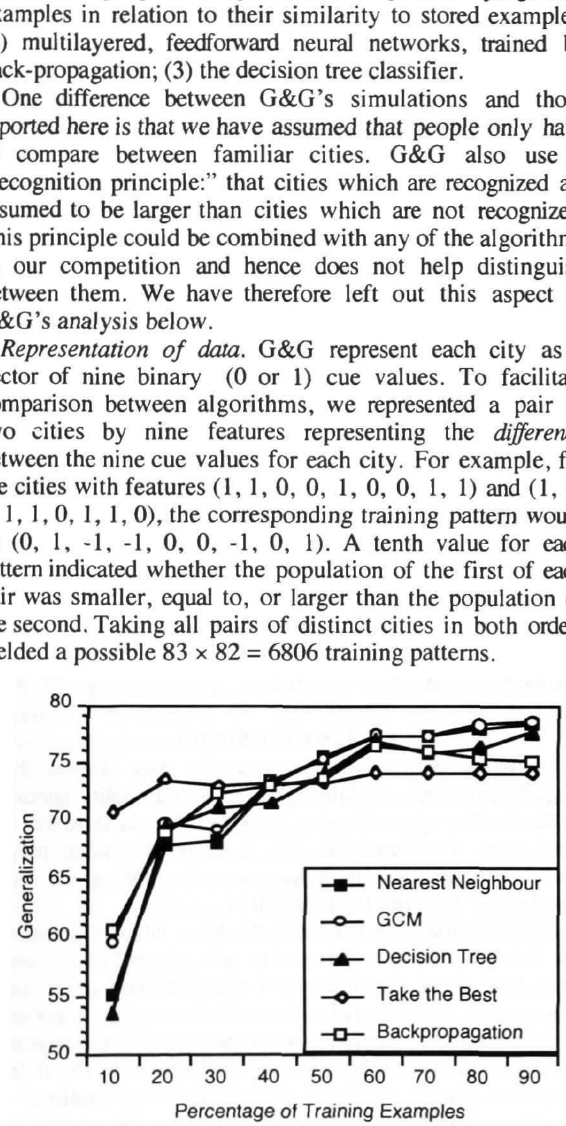

Figure 1. Results of the Competition. Percentage of correct inferences about the population of German cities as a function of the percentage of comparisons seen during training.

In order to capture the effects of limited knowledge, we trained each of the algorithms on a subset of the 6806 comparisons. In Figure 1, the percentage of training examples refers to the percentage of these comparisons presented during the training of each algorithm. The values shown in Figure 1 are for generalisation performance, for predicting the outcome of all 6806 comparisons. This approach allowed the algorithms to be assessed on an equal footing.

Take the Best. As described by G&G, we took a subset of the 6806 comparisons and treated them as training examples. The cue validity for each feature was calculated as the fraction of cases in which the feature was larger in value for the larger city divided by the total number of training examples. This gave the same cue validities as those calculated by G&G when the entire data set was used for training. Having found the cue validities, the test set was evaluated in the rank order specified by the cue validities, and the first feature in this order that discriminated between the two cities was taken to be the model's answer.

Nearest neighbor. The response of the model on test examples was that of the nearest neighbour in the nine dimensional space of training examples, using a Euclidean distance metric.

Generalised Context Model. The Generalised Context Model (Nosofsky, 1990) is similar to Nearest Neighbor, but the response is determined by all training examples, rather than just the nearest neighbor, in proportion to their similarity to the test example. Specifically, this method uses a Euclidean distance metric, and the influence of each training example is a Gaussian function of its distance from the test example. Nosofsky's model has adjustable parameters concerning the relative weighting of each feature, and also allows bias terms for different response. For simplicity we did not include such parameters, and thus each feature was weighted equally, and there were no biases between responses. We included just one adjustable parameter, the standard deviation of the Gaussian, which was optimised straightforwardly by measuring generalisation score for many different values and choosing the best.

Feedforward connectionist network. We used a three-layer feedforward network with nine input units, two hidden units, and one output, trained using the backpropagation algorithm. The inputs were the difference patterns, and the output corresponded to the decision about which city is larger. The target values for the output were 0, 0.5, and 1, for smaller, equal to, and bigger, respectively. Weights were initialised to random values within the range (-.5,.5). The net was trained for 100 epochs (passes through each training sample), with a learning rate of 0.01, and a momentum of 0.9. The order of the training examples was randomised within each epoch. During test, output values less than 0.5 were classed as "smaller," and values greater than 0.5 were classed as bigger.

Decision trees: C4.5. The decision tree building algorithm C4.5 (Quinlan, 1993) was used to construct a decision tree on the basis of the nine feature training vectors, and then used to classify the comparison vectors in the test set. See Quinlan (1993) for a detailed description and source code for the C4.5 algorithm.

Results and discussion. The results shown in Figure 1 show similar levels and patterns of performance for all the algorithms tested. The only substantial difference between algorithms is that Take the Best is most successful with a small number of training examples, and least successful with a large number of training examples. The overall levels of performance between Take the Best and the other algorithms are very similar. Combined with G&G's observation that Take the Best had almost exactly the same performance profile as tallying, weighted tallying and multiple regression, this suggests that this population estimation task is a poor discriminator between algorithms. Thus, from the results of the competition, there seems no reason to favor Take the Best over standard and widely used learning methods from psychology and artificial intelligence. We now argue that Take the Best is implausible on grounds of generality, and that, despite G&G, there is no reason to favor it on grounds of speed.

Generality

Take the Best is a highly specialized algorithm, along two dimensions. First, it is specialized with respect to *domain*: it applies only to the very restricted class of problems, in which some magnitude must be compared between pairs of items, where those items are represented by binary features. Second, it is specialized because it makes strong assumptions about the *structure of the data* that it can solve successfully: Specifically, Take the Best will succeed only for problems in which individual features can be meaningfully be considered independently. The other algorithms in the competition are more general than Take the Best, on both dimensions. They have all been used across a wide range of *domains*, including the modelling of disparate psychological processes. They also make weaker assumptions about the structure of the data—specifically, they can integrate information in complex way.

These general algorithms are able to perform as well as Take the Best in the competition we have reported. According to the standard principles of scientific methodology (e.g., Howson & Urbach, 1989), other things being equal, more general algorithms, which might potentially be common to many other cognitive processes, should be preferred.

Speed

G&G discuss one feature of Take the Best, which may appear to provide grounds for favoring this algorithm over the general purpose algorithms we have considered: that Take the Best draws inferences faster than the other algorithms in their competition, "measured by the amount of information searched in memory" (p. 658).

We note simply that the appropriateness of this measure depends on assumptions about the architecture of the cognitive system. On a serial architecture, in which it may be presumed that information is searched in memory at a constant rate, Take the Best would be more rapid than, for example, multiple regression, or the neural network model and exemplar accounts we have considered here. But in a parallel architecture, speed of processing will not generally be related to the amount of information searched in memory, because large amounts of information can be searched in memory simultaneously. So, for example, both the learning and application of multiple regression can be implemented in parallel using a connectionist network with a single layer of connections. This implementation could operate very rapidly—in the time it takes to propagate activity across one layer of connections. Similarly the back-propagation account could also be rapidly implemented in parallel, in connectionist hardware. In the same way, an instance-based architecture, in which instances can be retrieved in parallel, the nearest neighbor and general context model algorithms would be the quickest.

Only by making specific assumptions about the cognitive architecture that runs an algorithm is it possible to usefully compare the speed of Take the Best against the alternatives we have discussed. There are extensive research programs aimed at establishing the viability of instance-based and connectionist architectures as general accounts of cognitive architecture (e.g., Kolodner, 1993; Rumelhart & McClelland, 1986). Although the success of these programmes is yet to be decided, we should not apply a measure of speed, such as amount of information searched in memory, which assumes that the appropriate standard by which simplicity is to be judged is given by a serial architecture.

In this section, we have argued that general purpose learning methods give comparable results to Take the Best on G&G's population estimation task, are preferable on grounds of generality, and equally plausible on grounds of speed. We conclude that the cognitive plausibility of Take the Best remains to be established.

Conclusions

In this paper, we have argued for two claims. First, cognitive processes exhibiting bounded rationality should be understood as approximating some rational standard that may need to be discovered. Second, a range of general purpose learning algorithms are at least as plausible as the highly specialized Take the Best algorithm.

We hope that the approach we have advocated regarding the relation between rationality and algorithmic accounts may have broad application in understanding cognition in psychology, animal behavior and the social sciences. It promises to reconcile rational and mechanistic constraints in a range of contexts where the debate focuses on the different emphasis placed on these constraints. Both rational and mechanistic factors are important, because the system under study is presumed only to approximate, perhaps quite accurately or perhaps very coarsely, a rational solution.

Within this framework, the debate between rationality-based versus mechanistic explanation becomes a matter of emphasis and degree, rather than a fundamental divide. We suggest that in any debate of this kind, there should be a methodological imperative to explore rationality-based explanations—only by doing so can the scope of this level of explanation be assessed; and we caution that rationality-based explanation cannot be abandoned wholesale, without losing the ability to explain why the system under study is adaptive or successful.

References

Anderson, J. R. (1990). *The adaptive character of thought*, Hillsdale, N.J.: Lawrence Erlbaum Associates.

Anderson, J. R. (1991a). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14, 471-517.

Chase, W. G. & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.) *Visual information processing*. New York: Academic Press.

Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Clarendon Press.

Elster, J. (Ed.) (1986). *Rational choice*. Oxford: Basil Blackwell.

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

Gigerenzer, G. Hoffrage, U. & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswickian theory of confidence. *Psychological Review*, 98, 506-528.

Goldstein, D. G. & Gigerenzer, G. (1996). Satisficing inference and the perks of ignorance. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 137-141). Mahwah, NJ: LEA.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. La Salle, IL: Open Court.

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

Kolodner, J. L. (1993). *Case-based reasoning*. San Mateo, CA: Morgan Kaufman.

Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Nosofsky, R. M. (1990). Relations between exemplar similarity and likelihood models of classification. *Journal of Mathematical Psychology*, 34, 393-418.

Oaksford, M. & Chater, N. (1991). Against logicist cognitive science. *Mind and Language*, 6, 1-38.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631.

Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning*, 1, 121-152.

Paris, J. (1992) *The uncertain reasoner's companion*. Cambridge: Cambridge University Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufman.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann: Los Altos.

Reiner, R. (1995). Arguments against the possibility of perfect rationality. *Minds and Machines*, 5, 373--389.

Rumelhart, D. E. & McClelland, J. L. and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Volumes 1 & 2. Cambridge, Mass: MIT Press.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99-118.