Chapter 9

Bayesian brains and cognitive mechanisms: harmony or dissonance?

Henry Brighton and Gerd Gigerenzer

Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany

Introduction

The complexity of the cognitive system makes powerful metaphors such as the probabilistic mind and the Bayesian brain appealing on the one hand, but limited on the other. The trick is to not only harness their productivity, but also recognize their limits. One problem confronting the notion of the probabilistic mind and the accompanying 'quiet probabilistic revolution' is the apparent intractability of rational probabilistic calculation (Chater *et al.*, 2006b, p. 293). Rational probabilistic models, however, are not typically interpreted as algorithmic or mechanistic theories but functional level theories used to establish connections between observed behavior, a rational principle of inductive inference, and the structure of the environment. These correspondences tell us when the cognitive system is performing well and, to varying degrees, are used to suggest that human behavior is consistent with rational principles of inductive inference. From an algorithmic standpoint, how should these empirical findings be interpreted?

The distinction between functional and algorithmic level theories has its roots in what is now termed the rational analysis of cognition, an adaptationist program which aims to understand the structure and function of the cognition system as an adaptive response to the challenges posed by the environment (Marr, 1982; Shepard, 1987; Anderson, 1990; Oaksford & Chater, 1998). While working on a purely functional level, the tractability problem is in one sense irrelevant given that no commitment is made to a mechanistic level interpretation, but in another sense, unsatisfactory. Indeed, a principle objective of the rational analysis of cognition is to narrow down candidate algorithmic level theories by establishing empirically determined performance criteria. If the grand prize in cognitive science is uncovering both why minds do what they do and how they do it, then the productivity and scope of the metaphor would ideally extend to the process level.

Can the notion of the probabilistic mind be seamlessly extended to the algorithmic level, or there exist unmovable barriers to reconciling rational probabilistic models with plausible mechanisms of mind? We will examine these questions by considering

In N. Chater & M. Oaksford (Eds.), The probabilistic mind: Prospects for Bayesian cognitive science.
New York: Oxford University Press. © 2008

the metaphor of the probabilistic mind from an alternative adaptationist perspective, and one that views much of human inductive inference as relying on an adaptive toolbox of simple heuristics (Gigerenzer *et al.*, 1999; Gigerenzer & Selten, 2001). Unlike the notion of the probabilistic mind, the metaphor of the adaptive toolbox is rooted to an algorithmic level hypothesis, which proposes that adaptive behavior, and inductive inference in particular, is in large part the result of an interaction between processing simplicity and ecological context. This view leads to the notion of ecological rationality. Here, the cognitive system is viewed as adapted to the relevant aspects of its environment to the extent that it achieves good enough solutions using the limited resources it has available, rather than attempting to find optimal ones. On this view, organisms do not optimize but satisfy (Simon, 1996), which makes the notion of adaptive success for the organism—its ecological rationality—inseparably tied to an algorithmic level analysis.

What barriers, if any, stand between a synthesis of the study of ecological rationality and functional level probabilistic models? First, we consider the role of rationality and optimality in framing the study of cognition, and examine how these concepts represent key points of divergence between the study of ecological rationality and rational analysis. Second, we examine the consequences of the intractability of optimal probabilistic calculation, and propose that the statistical problem known as the bias/variance dilemma arises as a consequence, and represents a significant and often overlooked dimension of the functional problem facing the cognitive system (Geman *et al.*, 1992). The bias/variance dilemma brings into focus a connection between estimation error, ecological context, and the properties of learning algorithms. Therefore, in addition, it has the potential to bridge functional level models, simple heuristics, and the notion of ecological rationality. Ultimately, the adaptationist perspective should encompass both functional and algorithmic level analyses. Our guiding concern will be the understanding how these two levels of analysis can be aligned.

The rational and the psychological

We will focus on the problem of inductive inference. Given some sequence of observations, an organism makes a successful inductive inference to the extent that it selects a hypothesis, which is *probable* (Tenenbaum & Griffiths, 2001a), *predictive* of future observations (Anderson, 1991b, p. 479), or one which leads to a *succinct recoding* of past observations (Chater & Vitányi, 2003). Thus, prediction, probability, and coding length are fundamentally related concepts which point to the essence of the rational problem of inductive inference (Li & Vitányi, 1997). Although the formal instantiation of these concepts can lead to slight inconsistencies, we will not consider these inconsistencies here (Vitányi & Li, 2000; Grünwald, 2005). Next, it is worth setting out what role rational principles can play when used to examine, evaluate, and ultimately frame the problem of inductive inference facing biological organisms.

Functional and algorithmic level explanations

Rational principles can be used to construct functional level models of cognition. When used in this way the objective is to understand to what extent human behavior coincides with rational expectation, and to what extent rational principles 'point to deep functional reasons why our minds work the way that they do' (Tenenbaum & Griffiths, 2001b, p. 776). This level of abstraction dispenses with the need to specify how, in mechanistic terms, data are processed to yield behavior. Viewing the organism as a black box, 'the structure of such a theory is concerned with the outside world rather than what is inside the head' (Anderson, 1991a, p. 410). In contrast, an algorithmic level model aims to provide a mechanistic account of how the organism processes data in order to address the task. Such a model is algorithmic in the sense that it describes the steps required to transform inputs to outputs such that these steps could plausibly be implemented on some form of computing machinery.

In part, the rational analysis of cognition seeks functional level models in order to alleviate some of the problems in arriving at mechanistic accounts: 'If we know that behavior is optimized to the structure of the environment and we also know what the optimal relationship is, then a constraint on mental mechanisms is that they must implement that optimal relationship' (Anderson, 1991b, p. 471). The idea is that mechanistic theorizing does not typically center on adaptationist assumptions, but rather aims to fit 'the facts at hand', those potentially second order effects which may arise as a consequence of the deeper and more concisely articulated problem of being adapted to the structure of the environment.

We will critically examine rational principles of inductive inference as appropriate concepts with which to explain adaptive cognition. Probabilistic notions of rationality and optimization, as analytic tools, should be uncontroversial. They are theory neutral. But as concepts used to characterize and explain cognition we question their validity. Without doubt, a range of opinion exists on the extent to which the metaphor of the probabilistic mind refers to a purely functional theory concerned with a behavioral perspective on cognition, a normative theory implying that deviation from rational expectation reflects irrationality or maladaptation, or to a deeper property impacting on how, in mechanistic terms, the cognitive system actually processes data. To examine these perspectives, it will prove useful to distinguish between the notion of the probabilistic mind in the broad sense and in the narrow sense.

The probabilistic mind in the broad sense

What we will refer to as the broad sense of the metaphor is the familiar one from rational analysis, where a strict separation between function and process is maintained. Here, a functional model, beyond setting behavioral constraints on candidate mechanisms, remains mute on how these outcomes are arrived at. This perspective is broad given that it leaves the door open to a wide range of possible mechanisms capable of producing the observed behavior. For example, when Tenenbaum and Griffiths (2001b) state that 'we do not assert that any of our statistical calculations are directly implemented, either consciously or unconsciously, in humans minds, but merely that they provide reasons why minds compute what they do' (p. 776), they are clearly adopting a strict analytic separation between functional level and process level explanations. The strength of this approach rests on the range of settings in which a close fit between a model of the environment, the rational principle, and the behavioral findings are established. These findings catalog instances of the cognitive system performing well, sometimes to the degree that the observed behavior is interpreted as optimal (Griffiths & Tenenbaum, 2006a).

When used in this way, a rational principle, such as Bayesian statistics, 'provides a principled framework for understanding human inductive successes and failures, by specifying exactly what a learner is and is not justified in concluding given certain assumptions about the environment' (Tenenbaum & Griffiths, 2001b, p. 776). Correspondences such as these, where human behavior coincides with rational expectation, represent important empirical findings because they indicate that the cognitive system has performed extremely well, under the assumption that the rational principle provides an appropriate notion of success. A tentative conclusion might then be that, to one degree or another, the mind behaves as if it were Bayesian. Knill and Pouget (2004), for instance, remark on the 'myriad ways in which human observers act as optimal Bayesian observers' (p. 712). The strength of this viewpoint—the implied ability of the cognitive system to 'act Bayesian'—rests on the range of settings in which this finding holds, and the degree to which they imply that the cognitive system is maladaptive, or irrational, in the cases when it fails.

The issues of failure and maladaptation are problematic. After all, poor rational performance in one context could reflect an extremely well measured trade off for good performance in another. For the organism, the most effective deployment of limited processing resources for addressing the problems posed by the environment may well result in such a trade off being made. Clearly, the separation of the functional problem from the processing problem involves an idealization. Although idealizations can be productive, the come at a price. After all, if a Bayesian rational analysis is interpreted as 'specifying exactly what a learner is and is not justified in concluding' (Tenenbaum & Griffiths, 2001b, p. 776), and human behavior deviates from what is justified as result of a processing trade off, should this response be considered as irrational, reflecting a maladaptation? For instance, if a slightly less probable hypothesis is chosen over a more probable one, to what extent would this reflect a failure if such a choice can be arrived at far quicker and by using significantly fewer processing resources?

From a strictly functional level perspective, this argument makes little sense since the notions of adaptation and rationality are typically separated from issues of processing, and hence resource usage. Although Anderson's original formulation of the rational analysis of cognition considered the role of processing limitations in constraining the optimal response function, and productive examples of such considerations exist in, for example, the rational analysis of memory (Schooler & Anderson, 1997; Schooler, 1998), the role of processing limitations are often neglected. For the problem of inductive inference the impact of processing limitations are arguably harder to integrate, and what Anderson (1991b) referred to as the 'true Achilles heel of the rationalist enterprise' (p. 473) has been largely sidestepped in recent work. In the discussion to come we will discuss how constraints on processing change the functional problem quite significantly. This is why, from our perspective, an appropriate notion of adaptive success and rationality for biological organisms should take into account the contributors to function other than mere outcomes, they should also consider how these outcomes are achieved, and therefore extend to the algorithmic level (Simon, 1996; Todd & Gigerenzer, 2003).

The probabilistic mind in the narrow sense

What we will refer to as the narrow sense of the metaphor of the probabilistic mind are those interpretations which, to varying degrees, make a projection from the functional level description to an algorithmic level explanation. Ranging from speculative proposals, which consider the 'in principle' possibility, to full theoretical projection, this stronger interpretation considers probabilistic calculation as a potentially valid algorithmic level concept. Current opinion varies on if and when such an extension is justified. Certain forms of low level perception are viewed as the kinds of information processing problems for which this extension is likely to be valid, giving rise to proposals such as the 'Bayesian coding hypothesis', the idea that probability distributions may be neurally coded and rationally processed (Knill & Pouget, 2004). For other forms of cognition, and especially higher-level cognitive tasks such as decision making, the question is treated as an open one in need of exploration (Chater *et al.*, 2006a). The fly in the ointment for the probabilistic mind in the narrow sense is the fact that rational calculation—the direct use of rational principles as processing principles—tends to be computationally intractable for anything but trivial problems. We will tackle this issue at greater length in the coming discussion, but in the interests of completeness, it is worth pointing out that the boundary between functional level and algorithmic level theories is, for us at least, not always clear.

For instance, when considering the role of compression, an interpretation of the rational principle of induction by minimum description length (MDL), Feldman (2003) argues that 'the neglect of complexity in concept learning has stemmed from the ascendancy of exemplar theories' (p. 227) and 'human learning involves a critical element of compression or complexity minimization that is not present in exemplar models' (p. 230). Such an argument contrasts exemplar models (an algorithmic level theory) with the minimization of coding length (a functional level rational principle), and casts doubt on the former as a result of not conducting the explicit rational calculation implied by the latter. In a given setting, the inferences made by the exemplar model could in principle be consistent with those suggested by the rational principle but, obviously, arrived at without any form of explicit complexity minimization. Unless one views the minimization of coding length as a valid processing principle, such a comparison appears questionable. Fass and Feldman (2003), in a similar vein, consider that 'while it is premature to conclude that humans construct anything like the two part code [...], it seems likely that they employ some closely related complexity minimization principle'. Such a view implies that, to one degree or another, the cognitive system is itself applying the rational principle (the two part code interpretation of MDL, or something close to it). At the very least, there appears to be an implicit belief that organisms have the ability to perform something approaching rational calculation as if 'the machinery of probability and information theory' existed (Movellan & Nelson, 2001, p. 691).

Summary: from tools to metaphors

Our distinction between the metaphor of the probabilistic mind in the broad sense and in the narrow sense is an attempt to mark out degrees of projection of the analytic tools of probability and information theory to theories of the cognitive system itself (Gigerenzer, 1991). We accept that such a coarse grained distinction will miss some subtleties, but the distinction remains an important one. The broad sense of the metaphor is in large part an issue of degrees of idealization, and the theoretical price one is willing to pay for abstracting from the algorithmic level. The narrow sense of the metaphor is more of a technical issue, which we consider next by contrasting our own view of an adaptive toolbox with some of the specific features of the rational analysis of cognition.

The optimal and the psychological

An essential step in the rational analysis of cognition is to find the optimal response function. But to what degree does the goal of understanding the cognitive system as being adapted to problems presented by the environment require the notion of optimality? Is optimality merely an analytic tool for setting a performance benchmark, or a processing assumption made in order to support the metaphor of the probabilistic mind? For the broad sense of the metaphor, 'the optimal behavior function is an explanatory tool, not part of an agent's cognitive equipment' (Chater *et al.*, 2003, p. 70). For the narrow sense of the metaphor the role of optimality is problematic, as optimization as a cognitive processing principle is questionable from a tractability perspective. Until now, we have separated the issues of rationality and optimality because rational principles can be used without invoking the notion of optimality during the analytic process, as a feature of metaphor, or as an assumed property of the organism.

Adaptationism without optimality

Rational principles are required in order to say anything of substance about the success of the organism, or any model of the organism. After all, some normatively justified metric is required in order to inform the task of gauging success. Gauging success, on the other hand, does not require knowledge of the optimal solution. Furthermore, using a rational principle to inform an adaptationist analysis does not mean that optimality or rationality, when used in a broader sense, will necessarily be productive concepts with which explain the function and internal workings of the organism. To illustrate the point, two uses of a rational principle need to be distinguished. Functional models in rational analyses define an optimal benchmark against which behavior is judged, and something to be approached to varying degrees of approximation by the organism. Here, the rational principle is used as an *absolute* measure of function. Although the principle itself may well be interpreted as an approximate measure of function, it nevertheless sets a benchmark, and is absolute in this sense. Rational principles can also be used as model selection criteria, where competing processes are evaluated on their ability to perform the task adequately. Here, the rational principle is used as a *relative* measure of function. The optimal solution does not need to be known when a rational principle is used in this way.

For example, it is standard practice in machine learning to use Bayesian statistics, the MDL principle, or cross validation to assess the functional performance of learning algorithms in the absence of knowledge of the optimal response (Hastie *et al.*, 2001). Our analysis of simple heuristics, and how they perform in comparison to models carrying out more intensive forms of processing, relies on the use of rational principles as model selection criteria. Processes are viewed as performing better or worse than each other, rather than optimally or sub-optimally. On finding that a simple heuristic outperforms several intensive forms of processing model in a given ecological context, we examine the degree to which the simple heuristic is ecologically rational, and view it as a potentially interesting instance of how processing simplicity can be used to exploit the structure of the environment. Because we view cognitive processes as satisficing processes, which seek good enough inferences rather than optimal ones, there is no need to invoke the notion of optimality in order to assess this hypothesis (see also Vicente & Wang, 1998 for a similar perspective). We have no objection to using knowledge of the optimal solution to inform this process, but for many problems that we consider determining the optimal response is infeasible (Martignon & Laskey, 1999).

Comparing this approach to the practices of rational analysis, Chater and Oaksford (1999, p. 59) view our approach as being 'at least in the spirit of an optimality approach'. To clarify this issue, we view knowledge of the optimal response—when it can be reliably identified—as potentially useful knowledge, but not necessarily a useful concept with which to characterize cognitive processing, and certainly not a requirement for carrying out an adaptationist analysis. Indeed, often the significant difficulties in deriving the optimal response can restrict the problems one considers to 'toy world' settings lacking the complexity of the natural contexts faced by the organism. Thus, the methodological priority of identifying the optimal response can be a hindrance, particularly as it presupposes full and certain knowledge of the problem being considered.

In contrast, Chater and Oaksford (1999), argue that 'the need to explain the success of cognition means that, even if they are currently unavailable, deriving optimal solutions will remain a desideratum. Using Marr's analogy, ignoring this step of a rational analysis would be like trying to understand why birds can fly without a theory of aerodynamics.' (p. 59). Trying to understand the cognitive system from an adaptationist perspective without some normatively justified metric of success informing the analytic process, we agree, is likely to obscure the question. In this respect, rational principles should inform the analytic process to the extent that they provide a convincing model of functional success. But the same cannot be said for the role of optimality in adaptationism, both in our approach and, as others have argued, in Marr's (Gilman, 1996).

Optimizing processes and the problem of computational intractability

If the objective is to describe the behavior of an organism by comparing it to an optimal benchmark, then optimality is used as analytic tool only, with no commitment to the possibility or likelihood of optimal processing being a viable proposition. In the absence of an algorithmic level theory an optimal benchmark will often be the only non-arbitrary reference point available. However, if adherence to the optimal solution is taken to reflect or imply an optimizing process, then one must seriously consider if such a process provides a psychologically plausible and computationally tractable solution to the problem. This is why we view theories that consider optimization as a viable process level concept as often involving questionable idealizations that are likely to obscure the essence of the problem. Yet, on the other hand, all processes are optimal given a sufficiently contrived and narrow processing context.

The study of algorithms is often the study of approximating methods. Artificial intelligence, for example, usually concerns itself with the study of problems for which the optimal solution is either intractable or uncomputable (Simon, 1956; Reddy, 1988; Russell & Norvig, 1995). For the problem of inductive inference, ideal forms of rational calculation are uncomputable (Solomonoff, 1964; Li & Vitányi, 1997; Hutter, 2005). In more restricted and realistic settings, inductive inference using Bayesian belief networks can quickly become intractable even when one relaxes the objective to approximate Bayesian reasoning (Cooper, 1990; Roth, 1995; Dagum & Luby, 1993). Statistical machine learning, which provides a significant source of insight and motivation for those examining probabilistic cognition, is chiefly the study of approximation and, as Bishop (2006) points out, 'for many models of practical interest, it will be infeasible to evaluate the posterior distribution or indeed compute expectations with respect to this distribution' (p. 461) and 'in such situations, we need to resort to approximation schemes' (p. 462). Thus, even when the inference problem is reduced to quite restricted settings, tractable algorithms capable of yielding optimal rational outcomes remain illusive. Identifying the optimal solution for a specific problem by analytic means is often beyond reach, a task, which is significantly less tricky than specifying a tractable algorithm capable of arriving at the optimal solution for such problems in general.

Dealing with error: the bias/variance dilemma

How can a functional model suggest certain kinds of mechanisms and not others? In order to help the induction problem—the problem of identifying process level theories which can explain the adaptive success of the cognitive system—assumptions about the algorithmic level are required. In order to provide traction on the induction problem, how theory-specific do these assumptions have to be? Rather than making specific assumptions about the algorithmic layer, the issue we turn to now points to how quite a general formalization of the processing problem can be used to narrow down the kinds of process capable of approaching the significant performance requirements set by rational analyses of cognition.

The assumptions we start with are very general ones concerning the anatomy of inductive processes, and how they can be viewed as performing search over a model class. In this setting, the statistical problem of the bias/variance dilemma can then be used to narrow down the kinds of search procedures and model classes that will be successful in certain contexts (Geman *et al.*, 1992). Ultimately, we will argue that the contexts of interest in cognition, where good performance from sparse data appears to be a hallmark, point to the reduction of variance, or equivalently, the objective of imposing stability on the learning map, as fundamental problems to be overcome by cognitive processes. Simple heuristics, the processing model on which our research program is based, will ultimately be shown to address this objective. By framing the problem in terms of the bias/variance dilemma, bounded rationality can be given a statistical interpretation and justification, which has previously been lacking (Brighton, 2007).

The anatomy of an inductive process

All inductive processes can be formalized as maps from sequences of observations to hypotheses drawn from a hypothesis space. An observation is a pair composed of an input and an output. A particular environment can be thought as a joint probability distribution on an observation space, such that the combination of the two determine how likely each observation will be. We will sometimes refer to this environmental setting in terms of a *target function*, where the target function defines the form of relationship between inputs and outputs occurring in the environment. The task of the learning algorithm is to process sequences of observations in order to induce a hypothesis. The hypothesis space of the algorithm can also be viewed as a model class. A model is simply a parameterized family of hypotheses, were each hypothesis is a fully specified conditional probability distribution. A model class represents the set of models that the algorithm induces over.

Function and the bias/variance dilemma

Organisms process sequences of observations, samples of the target functions governing the environment. Inductive inference is the task of identifying the systems of regularity that govern this environment, given only these samples. An organism well adapted to this task should not be judged solely on its ability to perform well on a single sample. For example, the Bayes optimal classifier—as a process—is optimal only in the sense that it optimal on average. Other processes will outperform it if the sampling assumptions are violated. When estimating the mean predictive accuracy over many samples of the target function, the mean error of the algorithm can always be decomposed into three terms:

$$\text{Error} = \text{Irreducible Error} + \text{Bias} + \text{Variance}. \quad * \, (1)$$

Irreducible error is noise, and sets an upper bound on the achievable predictive accuracy. The remaining error can be decomposed and controlled through the design of the learning algorithm. This decomposition results in two terms, *bias* and *variance* (see Geman *et al.*, 1992, Bishop, 1995, and Hastie *et al.*, 2001 for derivations and further discussion). Across samples, bias is the difference between the mean predictions of the algorithm and the target function. Variance is the expected squared deviation about this mean, and arises because different hypotheses are likely to be induced for different samples of the target function.

The bias/variance dilemma.

The potential for an algorithm to achieve low bias will depend on how well it can approximate the underlying target function. General purpose processing methods, such as the nearest neighbor classifier and decision tree inductive algorithms, excel at achieving low bias by inducing over model classes with little, if any, restrictions on the functional form of the models. Consequently, a serious problem stands in the way of these methods, and nonparametric[1] methods in general, as being adequate process models of inductive inference. When the training sample is small, in the sense that it provides sparse coverage of the observation space, there is likely to be a potentially significant variance component to the error. Generally speaking, the smaller the size of the training sample, the higher the variance.

To combat the problem of high variance, restrictions on the model class are needed in order to impose stability on the learning map. But clearly, by restricting the model class the method will then suffer from high bias for certain classes of target function. To achieve accurate predictions across samples requires that a process must strike a good balance between reducing bias and reducing variance. Whether or not a process achieves a good balance depends entirely on context. Without stating the class of target functions likely to be encountered, practically nothing can be said about how well a process will achieve this balance when data are limited. This problem is known as the bias/variance dilemma (Geman *et al.*, 1992). All inductive processes can be thought of as making a bet, not only on what kinds of target function the environment will present, but also the likely degree of exposure to these target functions.

Sparse exposure and the context of induction

When functioning in a natural environment, the bias/variance dilemma will pose a significant problem for the organism: complete exposure to the systems of regularity occurring in the environment is typically not possible, observations are limited and often costly, and inductive inference is most pressing when there is a need to generalize to unseen cases. Indeed, the remarkable effectiveness of the cognitive system is seen as remarkable precisely because good inferences appear to be made despite sparse exposure to the underlying regularities (Tenenbaum *et al.*, 2006; Griffiths & Tenenbaum, 2006b, p. 130).

Importantly, the notion of the bias/variance dilemma was originally motivated by the need to account for these phenomena, and align mechanistic accounts with studies of cognition which propose that 'the brain is a proof of existence of near-optimal methods that do not require prohibitively large training samples' (Geman *et al.*, 1992, p. 46). The chief conclusion arising from this work is that, from a processing perspective, 'off the shelf' nonparametric methods such as feed-forward neural networks, nearest neighbor methods, and decision tree induction algorithms, fail as adequate responses to the bias/variance dilemma when data are limited. Without customization, they induce over ostensibly unrestricted model classes, a perspective which 'teaches us all too little about how to solve difficult practical problems' and 'does not

[1] Nonparametric methods are those which make minimal assumptions about the functional form of the data generating model (Geman *et al.*, 1992; Bishop, 2006).

help us out of the bias/variance dilemma for finite-size training sets.' (Geman *et al.*, 1992, p. 45).

The bias/variance dilemma implies that general purpose learning, in natural contexts of sparse exposure, is unachievable in any meaningful sense because tractable processes will suffer from error, and the degree of error is likely to vary significantly depending on the content and size of the training sample. Furthermore, it narrows down the kinds of processing strategies capable of meeting levels of performance suggested by rational analyses, and stresses the need to understand the context of induction. More generally, as soon as an organism makes inferences from impoverished data, the variance component of error becomes critical, and one must, to one degree or another, abandon the objective of nonparametric inference. The substantive question now is how this can be done.

Ecologically rational processing

A decomposition of the inference task, and the cognitive system more generally, is often viewed as necessary on grounds of computational tractability (Barrett & Kurzban, 2006; Samuels, 2005) and biological plausibility (Gallistel, 2000). For instance, skepticism toward the tractability of global Bayesian updating can be partially alleviated by updating on a within module basis, leading to the idea that one can 'jettison the goal of being globally Bayesian and instead assume only that each module is Bayesian itself' (Kruschke, 2006, p. 681). But once the black box is opened, on what basis should its contents be organized? For the problem of inductive inference, the bias/variance dilemma suggests that processes induce over constrained model classes in order to impose stability on the learning map, and hence reduce variance. Therefore, decomposition is not merely driven by issues of tractability, but is perhaps more fundamentally driven by issues of function.

We now attempt to tie together the general form of the relationship between functional level analyses, simple heuristics, and the bias/variance dilemma in order to say something about an alignment between ecological focus and processing simplicity. Our hypothesis is that constraints on cognitive processing can align a process with the structure of the environment. An extreme, but nevertheless entirely plausible consequence of this hypothesis is that conducting less processing is just as likely to reduce variance than conducting more processing. This explains why heuristics 'work', adding a statistical interpretation for why the mind might 'operate via a set of heuristic tricks, rather than explicit probabilistic computations' (Chater *et al.*, 2006a, p. 290). Our objective now is to say something about how the retreat from the objective of general purpose nonparametric inference can proceed and be given cognitive-ecological guidance.

Simple heuristics for the bias/variance dilemma

To flesh out our argument we will briefly examine the simple heuristic Take The Best (Gigerenzer & Goldstein, 1996) but frame it in different terms than previously used (see Brighton, 2007, for details). Take The Best is a cognitive process model for making inductive inference on the paired comparison task, where the problem is to rank two objects on their unobserved criterion values. This is a specific form of supervised concept learning. Training observations are pairs of objects, along with feedback on which object scores higher on the criterion. In an induction phase, Take The Best orders the cues by their validity. In the decision phase, it searches for the first cue in the order that discriminates between the two objects, and uses this cue alone to make a prediction. Validity is naïve measurement of a single cue, and simply captures the accuracy of the inferences made by this cue alone when inferring the rank of objects. By referring to Take The Best as simple, we are referring to the fact that it ignores conditional dependencies between cues when selecting a hypothesis, and does not weigh and add cues when making inferences.

The performance and analysis of take the best.

Take The Best often outperforms linear regression models and other simple heuristics over a wide range of environmental contexts (Czerlinski *et al.*, 1999). Using a more reliable model selection criterion than that used by Chater *et al.* (2003), Brighton (2007) shows, contrary to their findings, that Take The Best frequently outperforms a range of neural network, decision tree induction, and exemplar models. In short, Take The Best provides a good illustration of how performing less processing can lead to improved performance in natural environments. Understanding the environmental conditions under which Take The Best, and other simple heuristics, can outperform more computationally intensive methods is the next question.

When viewed in terms of bias/variance dilemma previous work focusing on this question can be seen as identifying conditions for low bias. Conditions for low bias tell us when an algorithm has the ability to closely approximate the target function given a large enough training sample. For example, the non-compensatory environments, those which have rapid decay in cue validities, point to the cases when Take The Best will perform as well as a linear model (Martignon & Hoffrage, 1999, 2002; Katsikopoulos & Martignon, 2006). But matching the performance of another linear model under these conditions is only guaranteed when there is a sufficiently large training sample to saturate the observation space and, crucially, such arguments offer no explanation for the fact that Take The Best can *outperform* a number of linear and nonlinear models. In short, previous analyses of when and why simple heuristics perform well do not consider the very statistical property which confers the performance advantage (Brighton, 2007).

Context sensitive induction.

To frame the performance of Take The Best in terms of the bias/variance dilemma, we will consider two further processing models, and two environments which will elicit drastically different relative performance between the models. The first (natural) environment is the often-studied German city environment (Gigerenzer & Goldstein, 1996; Chater *et al.*, 2003). The second (synthetic) environment is an instance of the more general class of non-compensatory environments, where cue validities decrease rapidly as a function of their rank (Martignon & Hoffrage, 1999, 2002). The two further models we consider are the well known decision tree induction algorithm CART (Breiman, Friedman, Olshen, & Stone, 1994), and a variant of Take The Best (labeled here as *TTB.CV*) which carries out the additional computations required to assess conditional dependencies between cues, and then ranks cues by conditional validity (Martignon & Hoffrage, 1999).

These two methods reflect two useful points for comparison. First, the model class of Take The Best is nested with respect to the model class of CART, since Take The Best is itself a decision tree induction algorithm, inducing trees with restricted function form. Second, the model class of Take The Best is identical to that of TTB.CV. The two methods differ only in how they perform search in order to select the cue order. Now, Figure 9.1(a) plots the predictive accuracy of Take The Best and these two models as a function of sample size for the German city population environment.

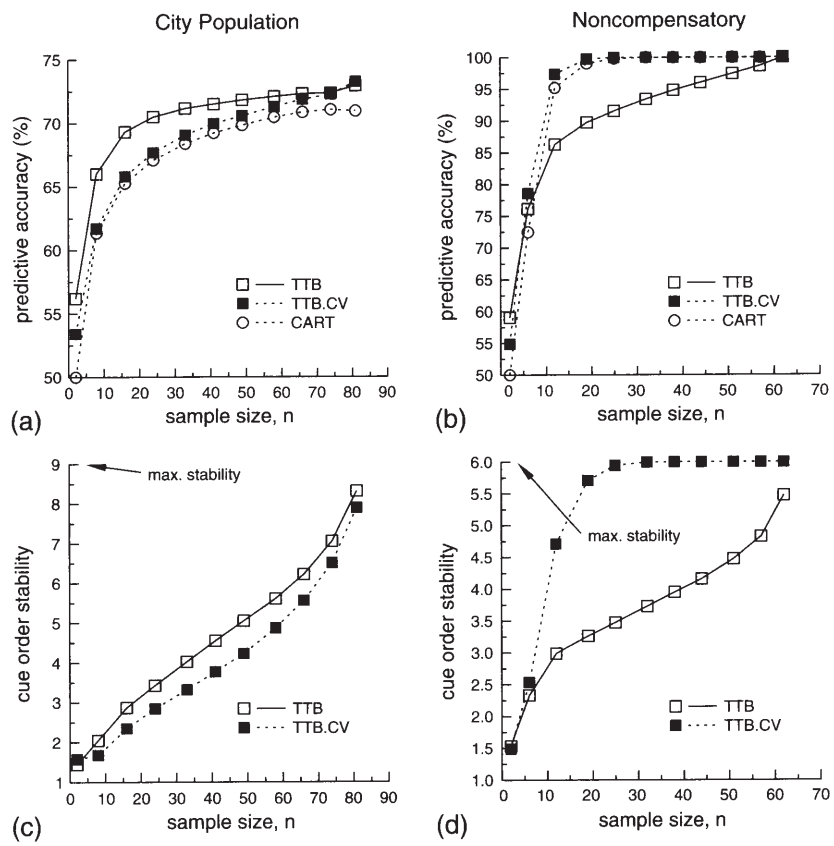

Fig. 9.1. A model comparison of Take The Best (labeled TTB), CART, and a variant of Take The Best which orders cues by conditional validity (labeled TTB.CV). Plot (a) compares the predictive accuracy of the models as function of sample size for the German city population task. Plot (b) compares the models in a synthetic non-compensatory environment. Plots (c) and (d) shows the average Levenschtein distance between induced cue orders as a function of sample size. Cue order stability predicts predictive accuracy very closely.

Predictive accuracy is estimated using cross-validation. Take The Best significantly outperforms both methods across the majority of sample sizes. Second, Figure 9.1(b) shows the same comparison for the non-compensatory environment. Now the other methods outperform Take The Best across the majority of sample sizes. These two environments illustrate how the performance of the process is determined not only by the environment, but also the size of the learning sample. Why is this?

Using search to control variance and stability.

When considering the contribution of bias and variance, Take The Best will tend to outperform a model with a richer models class, such as CART, as a result of reducing variance, since any function Take The Best can approximate, CART can too. But Take The Best is also able to outperform TTB.CV, which has an identical model class. This point illustrates that controlling variance is not simply a matter of placing restrictions on the model class, but can also arise as a consequence of restricting search (Mitchell, 1982; Domingos, 1999). To illustrate the point, the structural stability of the cue orders induced by Take The Best and TTB.CV can be measured directly, and their dependence on sample size and connection to accuracy clarified. For the German city population environment, Figure 9.1(c) shows how the structural stability of the cues orders—here measured as the mean Levenshtein (1966) distance[2] between induced cue orders—predicts almost exactly the relative difference in predictive accuracy of Take The Best and TTB.CV. Notice how CART and TTB.CV perform almost identically.

Figure 9.1(d) shows the same comparison for the non-compensatory environment. Again, stability reflects predictive accuracy: Take The Best performs well to the extent that it imposes stability on the learning map, and hence reduces variance (Turney, 1995; Poggio *et al.*, 2004). One way of thinking about the sensitivity of a cognitive process to the contents of particular samples of the environment is to view this instability as reflecting a failure to ignore accidental, unsystematic, and therefore unpredictive regularities. If a process ignores these accidental regularities and truly focuses on systematic ones, then differences in the content of samples of the target function should not matter too much. Crucially, the determining factor in imposing stability is not the model class itself, but how search is conducted over the model class. By performing less processing and ignoring conditional dependencies between cues, the ecological focus on the ability to achieve stability can be shifted from the synthetic non-compensatory environment (where TTB.CV excels) to the natural German city population environment (where Take The Best excels).

The implications of the bias/variance dilemma for processes and priors

When the variance component of error is the major source of error, the relationship between the properties of the process and the environment is not so clear. In this

[2] Levenshtein distance is the minimum number of additions, deletions, and substitutions required to transform one string into another. By interpreting cue indices as symbols, Levenshtein distance provides a distance measure between any two cue orders.

situation, assumptions made on the part of the process and their relationship to the structure of the environment can have a strong positive impact on performance, despite a clear mismatch between the two. For instance, many years of sustained interest in the naïve Bayes classifier is due to the fact that it can perform surprisingly well despite the assumptions made during processing being explicitly violated by the underlying target function (Domingos & Pazzani, 1997; Friedman, 1997; Kuncheva, 2006). With respect to a given process, findings such as these indicate that environmental conditions for low bias can be orthogonal to the conditions for low variance.

From a probabilistic perspective, given full knowledge of the regularities and probabilities governing the environment, and therefore a good model of the hypothesis space and the prior, Bayes optimal inference defines the rational outcome. On accepting that a tractable mechanistic instantiation of this process will be approximate, the variance component of error enters the picture and must be controlled in order to approach rational outcomes. Or, from an MDL perspective, the model in the model class, which reduces the stochastic complexity of the observed data to the greatest extent is the rational choice (Rissanen, 1997; Grünwald, 2005). Given that an exhaustive search through the model class will be infeasible, the use of heuristic search in order to approximate this choice is required. Again, a tractable mechanistic instantiation of the rational process will lead to variance when the performance of the process is measured for different samples of the target function (e.g., the mean compression rate).

For the organism, variance about this mean is important. It reflects the sensitivity of the inductive process to the particular contents of the samples. As soon as approximation is the name of the game, the bias/variance dilemma has to be tackled. The greater the sparsity of exposure to the environment, the more critical this problem becomes. And, this is clearly a statistical problem contributing to the functional success of the organism, since the inductive performance of the organism should not be highly sensitive to different potential encounters (different samples) of the environment. Given that processes are approximate, and not optimal, a significant part of the essence of inductive inference arises due to the realities of resource bounded computation (Simon, 1996). On this view, the issue of cognitive limitations, and how they may serve a functional role by helping to reduce variance, becomes a significant source of further questions. If the rational analysis of cognition and the associated development of the probabilistic view on cognition are to be reconciled with mechanistic accounts, then these issues need to be confronted.

For example, does the bias/variance dilemma imply that for different likely degrees of exposure to the environment, different hypothesis spaces and priors are required to control variance? Thus, on asking where the priors come from, does the bias/variance dilemma play a role? Furthermore, if an analysis of the structure of the environment can only be loosely connected to the assumptions required on the part of the process, does this represent a barrier to reconciling Bayesian analyses with process level accounts? More generally, machine learning and artificial intelligence are often viewed as a rich source of ideas for furthering the probabilistic view on cognition, but to what extent do these disciplines focus on problems with an essentially different character? Large samples, a focus on nonparametric inference, and little concern for cognitive plausibility may represent a counterproductive source of inspiration (Geman *et al.*, 1992). These are some of the questions that need to be addressed.

Summary and conclusion

The notion of probabilistic mind and the study on functional level rational models has been described as the 'the most exciting and revolutionary paradigm to hit cognitive science since connectionism' (Movellan & Nelson, 2001, p. 691). The benefits of this approach are often presented relative to the common practices of cognitive science, which suggest 'a ragbag of arbitrary mechanisms, with arbitrary performance limitations' (Chater & Oaksford, 1999, p. 63). It points to a dichotomy between purposive and mechanistic explanation, with the implication that one faces a choice between an adaptationist perspective relying on rational models abstracted from the algorithmic level, or a mechanistic one with limited prospects of informing purposive explanation (Anderson, 1991b; Chater & Oaksford, 1999). Although this dichotomy is to a certain extent an accurate reflection of current practices, is such an explicit distinction beneficial?

The study of ecological rationality and the adaptive use of simple heuristics is an adaptationist program, which in contrast to the proposed dichotomy, is rooted to an algorithmic level hypothesis. Rather than using the concepts of rationality and optimization to theorize about how the cognitive system might be adapted to its environment, the notion of ecological rationality addresses how good enough solutions can be found with limited processing resources. Here, the basis on which the cognitive system is judged to be adapted to its environment takes into account the specifics of processing, and how the limited resources available to the cognitive system are harnessed to achieve adaptive cognition. In this way, the objective of understanding adaptive cognition need not sacrifice our understanding of the realities of processing. Can these two orientations, which clearly share deep commonalities, be aligned? We have taken work on functional level analyses as providing a valuable insight: They indicate that human level performance and current approaches to cognitive processing and artificial intelligence do not match, in the sense that human performance sets an extremely high standard yet to be achieved reliably by computational models. Something beyond minor repair to existing processing metaphors may be required in order to bring them closer.

The bias/variance dilemma is all about the inevitabilities of error, and points to a fundamental connection between performance and ecological context. It suggests that cognitive mechanisms must effectively reduce variance in order to address the kind of inductive inference problems of interest to cognitive science, which typically involve considerable degrees of accuracy despite sparse exposure to the environment. We showed how the simple heuristic Take The Best confers function by exploiting the connection between processing simplicity, the structure of the environment, and variance reduction. In this sense, we have sought a connection between the rational probabilistic models of cognition and simple heuristics: simple heuristics offer a form of processing model that the cognitive system could rely on in order to reduce the variance component of error. Variance will inevitably arise given the extreme implausibility of optimal rational calculation and the need to generalize despite sparse exposure to the environment. Machine learning tends to address this problem by performing more processing (Schapire, 1990; Breiman, 1996). We suspect that the cognitive system does not have this option, and instead tackles the problem by performing less processing.

From an algorithmic perspective the mind achieves adaptive behavior to varying degrees depending on ecological context. Rational principles are to a certain extent required to substantiate this view, but as the centerpiece to a metaphor, or the guiding principle of a paradigm, we believe they obscure something of the essence of cognition. As with all metaphors, one pays some kind of a price. The price one pays clearly depends on the problem, and the kind of answers one is looking for. For us, the *why* question—why the cognitive system behaves as it does, and the *how* question—how it does it—should not be separated. Indeed, we would find it deeply surprising to find that evolution had overlooked the use of simple processing solutions as way of adapting the organism to the environment using limited resources. Examining the cognitive system using principles divorced from the impact of processing may, on this view, be a heavy price to pay.

Acknowledgments

We would like to thank Nick Chater, Lael Schooler, Henrik Olsson, Peter Juslin, and Nathan Berg for insightful discussion, criticism, and comments.

References

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

Anderson, J. R. (1991a). The adaptive nature of human categorization. *Psychological Review*, **98**(3), 409–429.

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

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychological Review*, **113**(3), 628–647.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford: Oxford University Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* New York: Springer.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, **24**(2), 123–140.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, P. J. (1994). *Classification and regression trees.* Belmont, CA: Wadsworth International Group.

Brighton, H. (2007). *Ecological rationality and the bias/variance dilemma.* (Manuscript under review)

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, **3**(2), 57–65.

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, **90**, 63–86.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006a). Probabilistic models of cognition: conceptual foundations. *Trends in Cognitive Sciences*, **10**(7), 287–291.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006b). Probabilistic models of cognition: Where next? *Trends in Cognitive Sciences*, **10**(7), 292–293.

Chater, N., & Vitányi, P. M. B. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, **7**(1), 19–22.

Cooper, G. F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, **42**, 393–405.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). Oxford: Oxford University Press.

Dagum, P., & Luby, M. (1993). Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, **60**, 141–153.

Domingos, P. (1999). The role of Occam's razor in knowledge discovery. *Data Mining and Knowledge Discovery*, **3**, 409–425.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, **29**, 103–130.

Fass, D., & Feldman, J. (2003). Categorization under complexity: A unified MDL account of human learning of regular and irregular categories. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 35–34). Cambridge, MA: MIT Press.

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, **12**(6), 227–232.

Friedman, J. H. (1997). On bias. variance, 0/1-loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*, **1**, 55–77.

Gallistel, C. R. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1179–1191). Cambridge, MA: MIT Press.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, **4**, 1–58.

Gigerenzer, G. (1991). From tools to theories: a heuristic of discovery in cognitive psychology. *Psychological Review*, **98**, 254–267.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, **103**(4), 650–669.

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: the adaptive toolbox*. Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gilman, D. (1996). Optimization and simplicity: computational vision and biological explanation. *Synthese*, **107**, 293–323.

Griffiths, T. L., & Tenenbaum, J. (2006b). Statistics and the Bayesian mind. *Significance*, **3**(3), 130–133.

Griffiths, T. L., & Tenenbaum, J. B. (2006a). Optimal predictions in everyday cognition. *Psychological Science*, **17**(9), 767–773.

Grünwald, P. (2005). Minimum description length tutorial. In P. Grünwald, I. J. Myung, & M. A. Pitt (Eds.), *Advances in minimum description length* (pp. 23–79). Cambridge: MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Hutter, M. (2005). *Universal artificial intelligence*. Berlin: Springer-Verlag.

Katsikopoulos, K. V., & Martignon, L. (2006). Naïve heuristics for paired comparison: Some results on their relative accuracy. *Journal of Mathematical Psychology*, **50**, 488–494.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, **27**(12), 712–719.

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospectve revaluation and highlighting. *Psychological Review*, **113**(4), 677–699.

Kuncheva, L. I. (2006). On the optimality of Naïve Bayes with dependent binary features. *Pattern Recognition Letters*, **27**, 830–837.

Levenshtein, V. I. (1966). Binary codes capable of detecting deletions, insertions and reversals. *Russian Physics Doklady*, **10**(8), 707–710.

Li, M., & Vitányi, P. M. B. (1997). *An introduction to Kolmogorov complexity and its applications*. New York: Springer-Verlag.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

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). Oxford: Oxford University Press.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparisons. *Theory and Decision*, **52**, 29–71.

Martignon, L., & Laskey, K. B. (1999). Bayesian benchmarks for fast and frugal heuristics. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 169–188). Oxford: Oxford University Press.

Mitchell, T. M. (1982). Generalization as search. *Artificial Intelligence*, **18**, 203–226.

Movellan, J. R., & Nelson, J. D. (2001). Probabilsitic functionalism: A unifying paradigm for the cognitive sciences. *Behavioral and Brain Science*, **24**(4), 690–692.

Oaksford, M., & Chater, N. (1998). *Rational models of cognition*. Oxford: Oxford University Press.

Poggio, T., Rifkin, R., Mukherjee, S., & Niyogi, P. (2004). General conditions for predictivity in learning theory. *Nature*, **428**, 419–422.

Reddy, R. (1988). AAAI presidential address: foundations and grand challenges of artificial intelligence. *AI Magazine*, **Winter 1988**, 9–21.

Rissanen, J. (1997). Stochastic complexity in learning. *Journal of Computer and Systems Sciences*, **55**, 89–95.

Roth, D. (1995). On the hardness of appproximate reasoning. *Artificial Intelligence*, **82**, 273–302.

Russell, S., & Norvig, P. (1995). *Artificial intelligence: a modern approach*. Englewood Cliffs, NJ: Prentice Hall.

Samuels, R. (2005). The complexity of cognition: tractibility arguments for massive modularity. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: structure and contents* (pp. 107–121). Oxford: Oxford University Press.

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, **5**(2), 197–227.

Schooler, L. J. (1998). Sorting out core memory processes. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 128–155). Oxford: Oxford University Press.

Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology*, **32**, 219–250.

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, **237**, 1317–1323.

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

Simon, H. A. (1996). *The sciences of the artificial* (*3rd ed.*). Cambridge, MA: MIT Press.

Solomonoff, R. J. (1964). A formal theory of inductive inference, part 1 and part 2. *Information and Control*, **7**, 1–22, 224–254.

Tenenbaum, J. B., & Griffiths, T. L. (2001a). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, **24**, 629–640.

Tenenbaum, J. B., & Griffiths, T. L. (2001b). Some specifics about generalization. *Behavioral and Brain Sciences*, **24**, 762–778.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, **10**(7), 309–318.

Todd, P. M., & Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Econonmic Psychology*, **24**, 143–165.

Turney, P. (1995). Bias and the quantification of stability. *Machine Learning*, **20**, 23–33.

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, **105**(1), 33–57.

Vitányi, P. M. B., & Li, M. (2000). Minimum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Transactions on Information Theory*, **46**(2), 446–464.