Psychological Bulletin
2012, Vol. 138, No. 3, 415–422

0033-2909/12/$12.00 DOI: 10.1037/a0026884

COMMENT

How the Bayesians Got Their Beliefs (and What Those Beliefs Actually Are): Comment on Bowers and Davis (2012)

Thomas L. Griffiths
University of California, Berkeley

Nick Chater
University of Warwick

Dennis Norris
Medical Research Council Cognition and Brain Sciences Unit,
Cambridge, England

Alexandre Pouget
University of Rochester

Bowers and Davis (2012) criticize Bayesian modelers for telling "just so" stories about cognition and neuroscience. Their criticisms are weakened by not giving an accurate characterization of the motivation behind Bayesian modeling or the ways in which Bayesian models are used and by not evaluating this theoretical framework against specific alternatives. We address these points by clarifying our beliefs about the goals and status of Bayesian models and by identifying what we view as the unique merits of the Bayesian approach.

Keywords: Bayesian inference, probabilistic models, theoretical frameworks, computational neuroscience

In their target article, Bowers and Davis (2012) present an extensive critique of Bayesian models of cognition. However, this critique makes general claims about Bayesian models based on careful selection of specific examples and has a singular focus on identifying the weaknesses of the Bayesian approach rather than considering its merits relative to other theoretical frameworks. As a consequence, the reader may come away with several misconceptions about the goals and status of Bayesian models of cognition, along with an overly pessimistic view of the prospects of this approach. In this comment, we attempt to correct these misconceptions and to identify what we see as the merits of the Bayesian approach over other theoretical frameworks for studying human cognition. We support our claims by appealing to many of the specific examples of Bayesian models cited by Bowers and Davis.

Misconceptions About Bayesian Models

The arguments presented by Bowers and Davis (2012) paint a picture of Bayesian models (and Bayesian modelers) that we do not believe is accurate. In this section, we identify and address six misconceptions that a reader of their article could obtain.

The Goal of Bayesian Modeling Is to Show That People Are Optimal

Bowers and Davis (2012) focus on the idea that Bayesian modelers seek to show that people perform optimally at particular tasks. However, this is rarely the goal of Bayesian modeling. Rather, Bayesian models typically aim to provide explanations of human behavior. Identifying optimal solutions to computational problems posed by the environment and comparing these optimal solutions to human behavior is the tool that is used to yield these explanations. That the solutions are optimal licenses a particular kind of explanation—an explanation of cognition in terms of function, known as a "teleological explanation"—allowing us to assert that the match between the solution and human behavior may be why people act the way that they do.[1]

Comparing human behavior to optimal solutions does not imply a belief that people are actually computing those optimal solutions. Computing optimal solutions in complex tasks is simply not a viable hypothesis about how the mind or brain works. For instance, we have about 1,000 olfactory receptors, and we can recognize 1,000–10,000 odors. Performing inference over these variables for an arbitrary pattern of olfactory receptor activation is computationally intractable. The same problem occurs in most realistic problems of inductive inference, such as object recognition, diagnosing diseases, and interpreting sentences. Rather than providing evidence that people are computing the optimal solutions, a cor-

Thomas L. Griffiths, Department of Psychology, University of California, Berkeley; Nick Chater, Warwick Business School, University of Warwick, Coventry, England; Dennis Norris, Medical Research Council Cognition and Brain Sciences Unit, Cambridge, England; Alexandre Pouget, Brain and Cognitive Sciences, University of Rochester.

Correspondence concerning this article should be addressed to Thomas L. Griffiths, Department of Psychology, University of California, 3210 Tolman Hall, Room 1650, Berkeley, CA 94720-1650. E-mail: tom_griffiths@berkeley.edu

[1] Note that the term *teleological explanation* does not imply that the evolution of neural mechanisms performing Bayesian computations will have been driven by the goal of performing those calculations. Rather, teleological explanation provides an account of why natural selection might favor one mechanism rather than another, as in the rest of biology.

respondence between these solutions and human behavior suggests that we should begin to explore approximate algorithms that can find decent solutions to these problems in reasonable time, as we discuss in detail below.

The authors distinguish between "Type 1" and "Type 2" answers to "why" questions, with Type 1 being the (apparently more common) claim that people are optimal in an unqualified sense, whereas Type 2 is the claim that people are optimal with respect to a set of assumptions. This distinction does not seem meaningful, as only Type 2 questions appear to be well posed. In other words, optimal behavior is not defined without a specification of the problem being solved, and the assumptions of the agent are a key part of any inductive problem. Because the assumption of optimality means the solution is fully determined by the problem, the content of any Bayesian model of cognition reduces to a set of claims about this problem.

Viewed in this light, a Bayesian model is just an empirical hypothesis, in the same way as models developed in other theoretical frameworks. Models can be evaluated against existing data, but should also be used to make new predictions that are tested in the laboratory. The main differences from other kinds of models are that (a) the content of the Bayesian model is a proposal about the problem that people are solving rather than a characterization of the mechanisms by which they might be solving it, and (b) the style of explanation is teleological rather than mechanistic.

This characterization of Bayesian models should make it clear that the Bayesian framework is a means of generating empirical hypotheses, rather than an assertion that people are optimal or that particular kinds of processes take place inside people's heads. This approach seems close to what the authors refer to as "methodological Bayesianism," which we believe is the dominant view among Bayesian modelers (as opposed to their "theoretical Bayesianism," which, as we note below, may not exist).

In rare cases Bayesian models are used to argue that people behave optimally on a specific task. For example, Griffiths and Tenenbaum (2006) used a Bayesian model to argue that people accurately incorporate information about the distributions of different quantities in forming predictions. We agree with Bowers and Davis (2012) that making an argument of this kind requires strong constraints on the assumptions that go into such models. However, we also believe that such constraints are commonly used in cases where Bayesian models are used to argue for optimality. For example, Griffiths and Tenenbaum collected empirical estimates of the appropriate prior distributions and derived their model predictions directly from these estimated distributions.

Bayesian Models Are Unfalsifiable and Overly Flexible

In evaluating claims about falsifiability, it is useful to distinguish between a *model* and a *theoretical framework*. A model is proposed to account for a specific phenomenon and makes specific assumptions in order to do so. A theoretical framework provides a general perspective and a set of tools for making models. For example, the Bayesian approach outlined above, connectionism (Rumelhart & McClelland, 1986), and symbolic cognitive architectures such as ACT–R (Anderson, 1993) are three theoretical frameworks, each of which can be used to define models of specific phenomena. Models are falsifiable, but frameworks are typically not. Rather, frameworks live or die based on their ability to generate models that are useful. For a detailed discussion of the role of models and frameworks in science, see Lakatos (1970).

We believe that specific Bayesian models are readily falsifiable (or, at least, as falsifiable as any empirical hypothesis—any hypothesis can be "saved" by suitable ad hoc adjustments to other aspects of the theory; e.g., Duhem, 1914/1954; Putnam, 1974). But the general Bayesian approach, as with any scientific framework, is not. Frameworks provide the tools for building specific models, which can be assessed against observation and experiment. Frameworks, of whatever kind, cannot be falsified directly. Rather, they can be productive (i.e., creating models with novel predictions, corroborated by experiment; generating new lines of theoretical and empirical inquiry), or they can be unproductive (i.e., creating models that are continually in need of ad hoc patching and that generate few corroborating novel predictions, or fresh theoretical insights).

The charge of Bowers and Davis (2012) that the Bayesian framework is unfalsifiable is, therefore, misconceived: The same charge could be applied to the germ theory of disease, quantum mechanics, or the theory of evolution by natural selection. Moreover, lack of falsifiability is a characteristic shared by all other theoretical frameworks that have been proposed for modeling cognition. The idea that human cognition emerges from the interaction of simple components (e.g., Rumelhart & McClelland, 1986) does not yield falsifiable predictions; the same is true for the proposal that human cognition is symbolic computation (e.g., Newell & Simon, 1976). We *would* be worried about a lack of falsifiability if we viewed the Bayesian framework as making the unconditional assertion that people are optimal, but it should be clear from the preceding discussion that this is not the case. It may, however, explain some of the concerns of Bowers and Davis.

Turning to models rather than frameworks, Bowers and Davis (2012) argue that Bayesian models are unfalsifiable because of their flexibility in making different assumptions about priors, likelihoods, and utility functions. We agree that there are potentially many degrees of freedom in Bayesian models but disagree that there are more degrees of freedom than for other kinds of models. Connectionist models have a great deal of flexibility in the choice of architecture, learning algorithm, initialization, and training set. Symbolic cognitive architectures have potentially infinite degrees of freedom in the specification of production rules and a slightly more constrained set of degrees of freedom in the mechanisms used to select productions. As in other modeling frameworks, these degrees of freedom are constrained through the plausibility of different assumptions and the practice of evaluating models by running experiments to test novel predictions. Being careful about degrees of freedom and using appropriate procedures for comparing and testing models are important things to keep in mind for all forms of computational modeling; they do not constitute a problem that is specific to Bayesian models.

Part of the reason that the degrees of freedom of Bayesian models might seem salient is that they are unusually transparent. Using a Bayesian model requires declaring what the priors, likelihoods, and utility functions involved might be. Other modeling approaches implicitly have analogues of these degrees of freedom (e.g., the learning algorithm used in a neural network can be interpreted as a prior distribution on the weights; Mackay, 1995), but the assumptions that are made about their values are less apparent. Unprincipled assumptions in other models are rarely the focus of explicit discussion. For example, connectionist models of word recognition generally set resting activations proportional to log frequency (e.g., Davis, 2010) or present words in proportion to the log of their frequency (Seidenberg & McClelland, 1989). Why choose a log function? The data show that ease of word recognition is generally inversely related to log frequency, so modelers choose a log function because it fits the data. Because the choice of function does not follow from any theoretical principles, modelers are free to choose any function they like. In contrast, a model like the Bayesian reader (Norris, 2006) explains the log frequency on the basis that the prior probability of a word can be approximated by the probability of the word as derived from word frequency counts.

Bayesian Modelers Believe That "Bayesian Processes" Underlie Human Cognition

Bowers and Davis (2012) introduce the notion of a "theoretical Bayesian" as one who believes "that the mind carries out or approximates Bayesian computations at the algorithmic level, in some unspecified way" (p. 393). This is a broad definition, since anybody who believes that people behave in a way that is consistent with Bayesian inference on a given task must believe that the mind is somehow approximating the Bayesian solution. The position is clarified through the definition of a "Bayesian algorithm" as one that must "(a) store priors in the forms of probability distributions, (b) compute estimates of likelihoods based on incoming data, (c) multiply these probability functions, and (d) multiply priors and likelihoods for at least some alternative hypotheses" (p. 393). Under this definition, we suspect theoretical Bayesians (like the "Bayesian fundamentalists" of Jones & Love, 2011) may not exist (Chater et al., 2011).

Most Bayesian models of cognition are defined at Marr's (1982) "computational level," characterizing the problem people are solving and its ideal solution. Such models make no direct claims about cognitive processes—what Marr termed the "algorithmic level." To use Marr's analogy, a computational level analysis plays a role in explaining cognition similar to that played by a mathematical theory of aerodynamics in explaining bird flight. The theory of aerodynamics says nothing about the anatomical mechanisms of bone and muscle that support flight other than that they must produce a solution with particular properties. When explaining cognition, it is easier to confuse levels of analysis, given that Bayes's rule can be viewed as a procedure for updating beliefs as well as a mathematical solution for what those beliefs should be. Nonetheless, most Bayesian models appeal to the mathematical solution rather than the algorithmic procedure.

Recently, advocates of Bayesian models have begun a more detailed exploration of how computational- and algorithmic-level models might relate (e.g., Shi, Griffiths, Feldman, & Sanborn, 2010; Vul, Goodman, Griffiths, & Tenenbaum, 2009). These approaches take the main constraint produced by successful computational-level analyses to be the idea that people must somehow be approximating Bayesian inference, and explore algorithms that have this property. However, these algorithms need not bear a direct resemblance to Bayesian inference or satisfy Points a–d in the definition above. For example, Shi et al. (2010) showed that certain classes of Bayesian models can be approximated with an exemplar model, a traditional form of psychological process model. The idea behind this approximation is that people store examples of past events in memory, which act like samples from the prior, and then activate these stored exemplars based on their similarity to observed data, which acts like the likelihood function. Priors and likelihood functions appear in proving that this algorithm approximates Bayesian inference but disappear into familiar psychological notions of memory and similarity in defining the algorithm itself. Although this is just one example, we believe that there need be nothing intrinsically "Bayesian" about algorithms that approximate Bayesian inference.

The attractiveness of exploring approximate algorithms as hypotheses about cognitive processes is that they clarify what is being approximated and the nature of the approximation. The resulting models can be viewed as "heuristics," but ones that we understand and can connect directly to optimal solutions, as opposed to the "bag of tricks" approach that is sometimes advocated as an alternative to Bayesian models (and by Bowers & Davis, 2012). We see the combination of optimal solutions and efficient approximation algorithms as a powerful set of tools for analyzing how people might solve intractable computational problems.

Bayesian Models Are Never Compared to Other Approaches

One of the most serious charges of Bowers and Davis (2012) is that Bayesian modelers fail in their scientific duty to compare their models to competing explanations. We believe that there is, in some contexts, an argument as to why such an approach might be scientifically reasonable, but we also believe that the charge is largely false (and that this can be illustrated by some of the articles cited by Bowers and Davis).

As noted above, Bayesian models of cognition are typically defined at Marr's (1982) computational level. As such, it is not clear that they are in competition with accounts proposed at other levels of analysis, making it less relevant to compare to these accounts. To return to the example of bird flight, we can imagine a physicist who has devised a model of bird flight based on aerodynamics and a biologist who has devised a model based on the properties of the muscles and bones in a bird's wing. Both of these models make predictions about how birds fly. However, it is not obvious that the models should be compared in their fit or that one model should be rejected in favor of the other if it fits better.[2] The two models can both be valid, giving a characterization of the phenomenon at different levels of analysis; they are not in competition with one each other. However, the physicist's model should be compared with other functional analyses, and the biologist's model should be held up against other mechanistic models.

This argument suggests a reason why it might be appropriate not to compare Bayesian models with other models that are framed in terms of cognitive or neural processes. However, in practice such comparisons are extremely common. To take some examples from the articles cited by Bowers and Davis (2012), Norris (2006)

[2] In fact, it is possible to argue that one might expect the mechanistic account to provide a better fit regardless. A mechanistic account that actually characterizes the mechanisms correctly should fit better than the correct functional account, since the functional account will not be able to capture deviations due to the constraints provided by the mechanisms.

compared his Bayesian model to a number of non-Bayesian theories; Lewandowsky, Griffiths, and Kalish (2009) were explicitly motivated by a comparison with heuristic accounts; Oaksford and Chater (1994) and Feldman, Griffiths, and Morgan (2009) identified predictions that discriminated their account from other non-Bayesian explanations; and Ma, Navalpakkam, Beck, van den Berg, and Pouget (2011) used model comparison to test their Bayesian models against seven non-Bayesian alternatives.

Probabilistic Population Codes Require Noisy Neurons

Bowers and Davis (2012) are remarkably assertive when it comes to evaluating the neural evidence for probabilistic inference. As far as they are concerned, there is no such evidence. To evaluate this claim, we first need to define what would constitute a neural Bayesian theory. As we discussed earlier, optimality is not the key concept. Instead, the central questions are as follows: Do neurons encode single values, or do they also encode information about the certainty of the stimulus? Can they represent probability distributions (or likelihood functions) even when they have multiple peaks? Moreover, do neural computations take into account these probability distributions? These are fundamental questions, particularly because almost all theories of neural computation proposed in the last 30 years have assumed that neurons encode single values.

We suspect that, even with this broader definition, Bowers and Davis (2012) would continue to assert that there is no evidence for probabilistic encoding in neural circuits. Indeed, they argue that one of the main theories of how neurons encode probability distributions (the probabilistic population code idea of Ma, Beck, Latham, & Pouget, 2006) is based on the assumption that neurons are noisy, which is problematic given that several laboratories have shown that neurons are a lot less noisy than previously suspected and may in fact be very reliable.

If one were to take the characterization of probabilistic population codes offered by Bowers and Davis (2012) at face value, a lack of variability in neurons would seem damning. However, contrary to what Bowers and Davis suggest, this theory is not based on the idea that neurons are stochastic devices and that internal noise is what allows them to encode probability distributions. Instead, following Marr (1982) and others, Ma et al. (2006) proposed that uncertainty comes from the nature of the computational problems faced by the brain: Sensory measurements are not sufficient to specify the value of the latent variables (such as the direction of motion of objects or their identity) with absolute certainty. For instance, recovering the three-dimensional structure of an image from random dot stereograms is an ill-posed problem, with an infinite number of solutions, even if the images themselves are noiseless. Importantly, this external uncertainty implies that the measurements must vary from trial to trial, not because of noise in the sensors, but because of variability in the world (e.g., there are infinitely many images corresponding to the concept of, say, "car"). This in turn generates variability in the response of the neurons. Such variability is not under the control of the nervous system (because it is not generated by the neurons themselves), but the nervous system gets to decide how to parameterize this uncertainty. The probabilistic population code described by Ma et al. provides a method for uncovering the parameterization used in the brain, based on the distribution of neural responses conditioned on the latent variable (e.g., the variability in middle temporal neurons for a variety of objects moving, say, rightward).

Is this theory proven wrong if neurons turn out to be nearly deterministic devices? The answer is clearly no. It is critical not to confuse variability due to the external world, which is the focus of Ma et al. (2006), with variability due to internal neuronal noise. Whether neurons are stochastic is irrelevant to the probabilistic population code theory. There are plenty of other ways to test this theory, and the results so far are encouraging: Neural variability closely follows the distribution used by Ma et al. (Graf, Kohn, Jazayeri, & Movshon, 2011), and neurons combine their inputs across sensory modalities (Fetsch, Pouget, DeAngelis, & Angelaki, in press) and across time (Beck et al. 2008) in a way that takes into account the reliability of the encoded signals, as predicted by the probabilistic approach.

Bayesian Modelers Do Not Believe Constraints From Biology and Evolution Are Important

One of the most puzzling assertions of Bowers and Davis (2012) is that the emphasis on computational-level analyses represented by Bayesian models of cognition implies "of course . . . that the findings from other domains (e.g., biology and evolution) will play a relatively minor role in constraining theories in psychology" (p. 406). This is by no means an obvious conclusion. Many Bayesian modelers share with Marr (1982) the belief that contributions to understanding the human mind are going to come from all three levels of analysis, even as they share with him the conviction that the computational level provides the most effective place to start understanding feats of inductive inference such as visual perception, understanding language, and learning causal relationships. The "function first" strategy highlighted by Griffiths, Chater, Kemp, Perfors, and Tenenbaum (2010) is not a function-only strategy, and it is not the only strategy that is likely to ultimately succeed. There are fundamental questions about the mind and brain that can only be answered at other levels of analysis, and the insights yielded by biology and evolution are going to play a key role in developing an integrated theory at all these levels.

The Merits of the Bayesian Approach

Readers of Bowers and Davis (2012) might come away wondering why anybody would bother to make Bayesian models of cognition, as these models seem to provide few unique insights into the nature of the mind. In this section, we highlight some of the factors that we see as merits of the Bayesian approach when compared to other theoretical frameworks.

Universal Laws

The teleological explanations yielded by Bayesian models of cognition are valuable not just because they satisfy our desire to answer why questions, but because they provide the foundation for universal laws of cognition—principles that we expect to hold true for intelligent organisms of any kind, anywhere in the universe. Shepard (1987) made a classic argument for the virtue of such laws and provided his candidate for the first such law—the universal law of generalization. This argument rested on a Bayesian analysis of the problem of generalization (see Tenenbaum & Griffiths, 2001, for details).

Deriving General Predictions

Bayesian models of cognition cast many different phenomena in a single framework, and Bayesian inference provides a general account of learning and memory that provides the basis for many specific models. As a consequence, predictions that result from mathematical analyses of models based on Bayesian inference potentially apply across a wide range of domains. A relevant example taken from the articles criticized by Bowers and Davis (2012) is the analysis of serial reproduction given by Xu and Griffiths (2010). Contrary to the discussion of this article by Bowers and Davis, the goal of this analysis was not to show that a rational account could be given of biases in reconstruction from memory (Huttenlocher, Hedges, & Vevea, 2000, had already provided such an account). Rather, the goal was to test a general prediction about how information should change when passed from person to person that resulted from assuming that human learning and memory could be modeled as Bayesian inference. This prediction—that information should change through transmission to come to match people's prior distributions (Griffiths & Kalish, 2007)—is expressed at a level of generality where it can be tested in any domain, with any kind of stimuli. This generality results from having a general account of learning and memory, in the form of Bayesian inference, and is valuable because it means that the same simple result can provide an explanation for aspects of language evolution (Griffiths & Kalish, 2007) as well as a new method for estimating people's prior distributions (Lewandowsky et al., 2009).

Understanding Why Particular Mechanisms Work

Bowers and Davis (2012) object in several places that phenomena captured by Bayesian models were already explained by existing mechanisms or might easily be explained by appealing to intuitive mechanisms. For example, the phenomenon of reconstruction from memory explored by Xu and Griffiths (2010) and Huttenlocher et al. (2000) might be explained "as long as memory of an event is biased toward preexisting knowledge" (p. 400), and the analysis of the perceptual magnet effect by Feldman et al. (2009) is criticized as being consistent with a variety of algorithmic models. These complaints might reflect a particular set of beliefs about the goals of cognitive science, which can be illustrated through a simple thought experiment.

Imagine, at some point in the future, that cognitive scientists have succeeded in identifying the cognitive and neural processes that underlie all aspects of human behavior. Is the task of cognitive science now complete? We suspect that Bowers and Davis (2012) might say yes, but many Bayesian modelers would say no. Left open are questions about why these are the particular mechanisms that are used, whether there is a simple unifying theory that can explain them, and whether there are principles that can allow computers to behave similarly without instantiating the same cognitive and neural processes.

This thought experiment makes it clear how Bayesian models can be useful even in contexts where the psychological mechanisms are known. The Bayesian model of reconstruction from memory indicates not just that information from the past should be used, but exactly how it should be used. This can be captured by a simple mechanism, but the mechanism is now no longer arbitrary. The analysis of the perceptual magnet effect likewise explains why several previous models were able to capture this effect: They were approximating the optimal solution to the problem. To be fair, the concern that Bowers and Davis (2012) expressed was that this analysis placed few constraints on possible mechanisms. It actually imposes a strong constraint: We should expect mechanisms that approximate Bayesian inference to be able to produce the perceptual magnet effect. It just happens that many of the mechanisms that decades of research in cognitive science have converged on do possess this property—something that is hardly an accident.

Note, though, that the thought experiment is, in most contexts, very far from reality. It seems to us just as inconceivable that the mechanisms of cognition can be understood in the absence of an understanding of their function as that we might be able to unravel the functioning of a pocket calculator without any notion that it is doing arithmetic. Given the large role that uncertainty plays in the problems that people need to solve, we believe that Bayesian analysis will be play a central role in specifying the function of different aspects of human cognition.

Identifying Commonalities Between Different Mechanistic Accounts

Focusing on the abstract computational problems underlying human cognition can sometimes yield insights that are blurred when thinking purely in terms of mechanisms. In particular, this approach can make it possible to recognize the commonalities between existing theories, and to use this as the basis for identifying new theoretical approaches. One salient example is the analysis of category learning given by Ashby and Alfonso-Reese (1995), who showed that category learning could be analyzed as a problem of density estimation and that popular psychological process models such as exemplar and prototype models corresponded directly to different strategies for density estimation that had been explored by statisticians. This insight identifies a commonality between these different approaches and immediately provides access to a theoretical literature about the circumstances where one approach or the other might be expected to be more successful. Another example is the analysis of causal induction given by Griffiths and Tenenbaum (2005), who used causal graphical models to show that previous psychological models of causal induction had focused on the problem of estimating the strength of a causal relationship, neglecting the structural question of whether a relationship existed. Considering this question led to a novel model of causal induction that performed well in cases that were problematic for previous models.[3]

Coherence

Bowers and Davis (2012) substantially underplay the centrality of coherence in cognition. When, say, judging the depth of a

[3] We note that this article provides another example of a case where a Bayesian model was extensively compared with competing accounts.

nearby tabletop, the cognitive system is not solving a singular problem, using a set of specialized tricks for that purpose. Rather, it builds a representation of the layout of the scene, of which the tabletop is a part, specifying the depths of locations and orientations of table, wall, desk, the viewer's own body, as well as the location of the vertical, origin of light sources, shadows, and so on. The components of this representation are not computed separately but interact in complex ways (that a manuscript occludes the surface of the desk indicates that it is closer than the desk; the orientation of the water surface in a drinking glass indicates that the desk is tilted; and so on). In short, the process of building a model of the world from a sensory stimulus cannot result from the operation of independent mechanisms drawn from a "bag of tricks" (Ramachandran, 1990).

Enforcing coherence constraints requires being able to determine which degrees of belief about particular aspects of the scene are (or are not) compatible. The rules of probability are precisely coherence constraints on degrees of belief. Indeed, a variety of formal arguments (e.g., Savage, 1954) show that violation of the laws of probability will lead to incoherent beliefs and actions. Thus, once we allow that the perception and cognition are concerned not merely with individual judgments, but building coherent models of the world, then a Bayesian analysis becomes close to inevitable. This approach is, indeed, widely applied in computer vision (e.g., Zhu, Chen, & Yuille, 2009).[4]

Coherence constraints are equally crucial in language understanding (e.g., the interpretation of each part of the speech signal, as well as background knowledge, will affect the interpretation of the rest). Similarly, coherence constraints are crucial in general knowledge. Suppose it were the case that we have special-purpose mechanisms for answering questions such as whether Berlin or Hamburg is the larger city (e.g., Gigerenzer & Goldstein, 1996). If so, we should presumably have, equally, special-purpose mechanisms for determining whether Berlin, Hamburg, etc., are capital cities, whether these cities have soccer teams, whether they are in the former East or West Germany, or, indeed, whether capital cities are generally larger than other cities, and so on. Yet, on the face of it, such special-purpose mechanisms seem liable to lead to incoherence. Why should we not simultaneously believe, for example, that Berlin is larger than Hamburg (because it is a capital city), but that capital cities are generally smaller than other cities? Or suppose that we employ a different special-purpose algorithm to estimate the number of inhabitants of a city (e.g., Hertwig, Hoffrage, & Martignon, 1999). Might this not lead to contradictory conclusions concerning the relative size of two cities? A key attraction of Bayesian methods is that they provide a principled way of establishing coherence constraints and avoiding such inferential chaos.

Similarly, we suggest that an important consideration for cognitive science is that the brain is not concerned merely with answering independent "one-off" questions (for which special-purpose mechanism might be applicable), but with building a coherent general-purpose model of the external world that can be used to deal with a vast range of (potentially unforeseen) questions and challenges. Bayesian models provide a way to explain how our rich and (at least locally and partially) coherent probabilistic knowledge is used in, and the sophisticated and flexible inferences we make about, the world.

Probability Matching as a Case Study

The one topic for which Bowers and Davis (2012) explicitly identify an alternative to Bayesian models is in their treatment of probability matching, which they view as posing a challenge to accounts that assume a rational relationship between beliefs and behavior. They correctly point out that when presented with multiple alternatives with different probabilities of producing a reward, a rational agent should always choose the alternative with the highest probability. Instead, people choose alternatives with frequencies proportional to their probabilities. Bowers and Davis propose a mechanism to explain why this might occur in a sequence of choices, based on a neural network that implements a "win–stay, lose–shift" strategy of repeating a choice until it fails to produce a reward.

Probability matching is an interesting choice of example, as it is one of the phenomena (together with order effects; Kruschke, 2006) that have played a key role in explorations of possible mechanisms for approximating Bayesian inference. In particular, probability matching is what one might expect if people sample from a probability distribution over alternatives. Many Bayesian models make the assumption that this is how people select responses, based on a long tradition of behavioral models of choice (e.g., Luce, 1959). However, these models reveal a more subtle pattern than simply matching the probabilities of rewarded outcomes: In many experiments, people seem to be probability matching to the posterior distribution produced by the Bayesian model. For example, in the data of Griffiths and Tenenbaum (2006), the distribution of people's predictions was similar to the posterior distribution produced by the Bayesian model. Goodman, Tenenbaum, Feldman, and Griffiths (2008) found the same pattern for the rules selected by participants in a category-learning task.

Probability matching to the posterior is a more challenging phenomenon to explain than probability matching to the frequency of reward. In particular, it requires a mechanism that approximates Bayesian inference. Vul et al. (2009) proposed an explanation for this phenomenon based on the idea that people might be making decisions based on a single sample from a posterior distribution. This can be shown to be a reasonable strategy when the value of a correct decision is low relative to the cost of generating samples. This account also addresses the move toward deterministically selecting the highest probability outcome that Bowers and Davis (2012) raise as an issue for analyses based on sampling, as greater computational resources or greater value for a correct decision should result in generating more samples and thus favor higher probability outcomes. A variety of simple cognitive processes can be used to obtain samples from posterior distributions, including the exemplar model approach of Shi et al. (2010) introduced above and an algorithm that is based on the win–stay, lose–shift principle (Bonawitz, Denison, Chen, Gopnik, & Griffiths, 2011). The sim-

[4] Coherence constraints can be embodied mechanistically in, for example, constraint-satisfaction algorithms, as used in both symbolic and connectionist models (e.g., Rumelhart & McClelland, 1986; Waltz, 1972). Such algorithms are implementations of (approximations to) Bayesian methods, rather than alternative approaches. Note that recent technical developments, such as Bayesian graphical models (e.g., Pearl, 1988, 2000), have provided a richer understanding of such algorithms than was previously available.

ilarity of this approach to that proposed by Bowers and Davis suggests that the gap between Bayesian and non-Bayesian models may not be as large as it might appear.

Conclusion

Bowers and Davis (2012) summarize their argument (somewhat ironically) through an application of Bayes's rule. In the same ironic vein, we believe that there are several reasons why it is inappropriate to appeal to Bayes's rule in this case. First, the hypothesis that people are optimal is not something that even the most fervent Bayesian believes. Its prior probability, and hence its posterior probability, should be zero. Second, the alternative hypotheses are not specified. Despite chastising Bayesian modelers for failing to compare to alternative accounts, Bowers and Davis do not identify specific frameworks to which the approach can be compared or attempt to evaluate the merits of these frameworks. Finally, and perhaps most importantly, the hypotheses being evaluated are not mutually exclusive, and the problem that we want to solve is not determining which of these hypotheses is true. Rather, the question is how we as scientists should organize our efforts. Different theoretical frameworks, such as Bayesian modeling, connectionism, and production systems, have different insights to offer about human cognition, distributed across different levels of analysis. A connectionist model and a Bayesian model of the same phenomenon can both provide valuable information—one about how the brain might solve a problem, the other about why this solution makes sense—and both could well be valid. The ultimate test of these different theoretical frameworks will be not whether they are true or false, but whether they are useful in leading us to new ideas about the mind and brain, and we believe that the Bayesian approach has already proven fruitful in this regard.

References

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology, 39,* 216–233. doi:10.1006/jmps.1995.1021

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., . . . Pouget, A. (2008). Bayesian decision making with probabilistic population codes. *Neuron, 60,* 1142–1152. doi:10.1016/j.neuron.2008.09.021

Bonawitz, E., Denison, S., Chen, A., Gopnik, A., & Griffiths, T. L. (2011). A simple sequential algorithm for approximating Bayesian inference. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2463–2468). Austin, TX: Cognitive Science Society.

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin, 138,* 389–414. doi:10.1037/a0026450

Chater, N., Goodman, N. D., Griffiths, T. L., Kemp, C., Oaksford, M., & Tenenbaum, J. B. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral and Brain Sciences, 34,* 194–196. doi:10.1017/S0140525X11000239

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117,* 713–758. doi:10.1037/a0019738

Duhem, P. (1954). *The aim and structure of physical theory.* Princeton, NJ: Princeton University Press. (Original work published 1914).

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review, 116,* 752–782. doi:10.1037/a0017196

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (in press). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience.* doi:10.1038/nn.2983

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669. doi:10.1037/0033-295X.103.4.650

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32,* 108–154. doi:10.1080/03640210701802071

Graf, A. B., Kohn, A., Jazayeri, M., & Movshon, J. A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature Neuroscience, 14,* 239–245. doi:10.1038/nn.2733

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences, 14,* 357–364. doi:10.1016/j.tics.2010.05.004

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science, 31,* 441–480. doi:10.1080/15326900701326576

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 354–384. doi:10.1016/j.cogpsych.2005.05.004

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17,* 767–773. doi:10.1111/j.1467-9280.2006.01780.x

Hertwig, R., Hoffrage, U., & Martignon, L. (1999). Quick estimation: Letting the environment do some of the work. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 209–234). New York, NY: Oxford University Press.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129,* 220–241. doi:10.1037/0096-3445.129.2.220

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences, 34,* 169–188. doi:10.1017/S0140525X10003134

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review, 113,* 677–699. doi:10.1037/0033-295X.113.4.677

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge, England: Cambridge University Press.

Lewandowsky, S., Griffiths, T. L., & Kalish, M. L. (2009). The wisdom of individuals: Exploring people's knowledge about everyday events using iterated learning. *Cognitive Science, 33,* 969–998. doi:10.1111/j.1551-6709.2009.01045.x

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York, NY: Wiley.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9,* 1432–1438. doi:10.1038/nn1790

Ma, W. J., Navalpakkam, V., Beck, J. M., van den Berg, R., & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience, 14,* 783–790. doi:10.1038/nn.2814

Mackay, D. J. C. (1995). Probable networks and plausible predictions—A review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems, 6,* 469–505.

Marr, D. (1982). *Vision.* Cambridge, MA: MIT Press.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19,* 113–126. doi:10.1145/360018.360022

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113,* 327–357. doi:10.1037/0033-295X.113.2.327

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631. doi:10.1037/0033-295X.101.4.608

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, England: Cambridge University Press.

Putnam, H. (1974). The "corroboration" of theories. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (Vol. 2). La Salle, IL: Open Court.

Ramachandran, V. (1990). Interactions between motion, depth, color and form: The utilitarian theory of perception. in C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 346–360). Cambridge, England: Cambridge University Press.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

Savage, L. J. (1954). *The foundations of statistics.* New York, NY: Wiley.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568. doi:10.1037/0033-295X.96.4.523

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323. doi:10.1126/science.3629243

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review, 17,* 443–464. doi:10.3758/PBR.17.4.443

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640. doi:10.1017/S0140525X01000061

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? Optimal decisions from very few samples. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 148–153). Austin, TX: Cognitive Science Society.

Waltz, D. L. (1972). *Generating semantic descriptions from drawings of scenes with shadows* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

Xu, J., & Griffiths, T. L. (2010). A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology, 60,* 107–126. doi:10.1016/j.cogpsych.2009.09.002

Zhu, L., Chen, Y., & Yuille, A. L. (2009). Unsupervised learning of probabilistic grammar–Markov models for object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31,* 114–128. doi:10.1109/TPAMI.2008.67

Received November 14, 2011
Revision received November 14, 2011
Accepted November 26, 2011 ■