Psychological Bulletin
2012, Vol. 138, No. 3, 423–426

0033-2909/12/$12.00 DOI: 10.1037/a0027750

REPLY

Is That What Bayesians Believe? Reply to Griffiths, Chater, Norris, and Pouget (2012)

Jeffrey S. Bowers
University of Bristol

Colin J. Davis
Royal Holloway University of London

Griffiths, Chater, Norris, and Pouget (2012) argue that we have misunderstood the Bayesian approach. In their view, it is rarely the case that researchers are making claims that performance in a given task is near optimal, and few, if any, researchers adopt the theoretical Bayesian perspective according to which the mind or brain is actually performing (or approximating) Bayesian computations. Rather, researchers are said to adopt something more akin to what we called the *methodological Bayesian* approach, according to which Bayesian models are statistical tools that allow researchers to provide teleological explanations of behavior. In our reply we argue that many Bayesian researchers often appear to be make claims regarding optimality, and often appear to be making claims regarding how the mind computes at algorithmic and implementational levels of descriptions. We agree that some Bayesian theorists adopt the methodological approach, but we question the value of this approach. If Bayesian theories in psychology and neuroscience are only designed to provide insights into teleological questions, we expect that many readers have misunderstood, and hence there is a pressing need to clarify what Bayesian theories of cognition are all about.

Keywords: Bayes, Bayesian, optimal, algorithm, cognition

Bayesian theories in psychology and neuroscience have generated a great deal of excitement in recent years. A topic search of Web of Knowledge lists 51 articles published in *Psychological Review* in response to the term *Bayesian*, and 35 of these were published between 2005 and 2011. In response to the same search term, 14 out of 17 articles from *Nature Neuroscience* and 22 out of 24 articles in *Current Biology* were published since 2005. In our view, this excitement is related to the many strong and surprising claims about how we achieve near-optimal performance in a wide range of tasks in all variety of domains, from low-level vision to memory, language, reasoning, and motor control.

However, Griffiths, Chater, Norris, and Pouget (2012) argue that we have misunderstood the Bayesian approach. In their view, it is rarely the case that researchers are making the claim that performance in a given task is near optimal, and few, if any, researchers adopt the *theoretical Bayesian* perspective, according to which the mind or brain is actually performing (or approximating) Bayesian computations. Rather, researchers are said to adopt something more akin to what we called the *methodological Bayesian* approach, according to which Bayesian models are statistical tools that can be used to constrain theories: If behavior can be captured with a Bayesian model, it allows researchers to make teleological explanations about *why* people act the way that they do.

We find these claims regarding optimality and the theoretical Bayesian perspective surprising and remain unconvinced regarding the merits of the methodological Bayesian approach. Below we briefly consider these and related issues.

Bayesian Theories and Optimality

The claim that Bayesian researchers seek to show that people are optimal in a given task is hard to reconcile with the many statements in the literature, such as the quotes at the start of our target article (Bowers & Davis, 2012). One only has to look at the titles of many articles to understand where we and others have formed the impression that advocates of Bayesian models are highlighting the optimality of human performance in various domains (e.g., Ernst & Banks, 2002; Faisal & Wolpert, 2009; Feldman, Griffiths, & Morgan, 2009; Norris, 2006). In any case, in order to even get started on a Bayesian analysis, it has to be assumed that we are (near) optimal. As Griffiths et al. (2012) put it, the fact that "solutions are optimal licenses a particular kind of explanation . . . known as a 'teleological explanation' " (p. 415). Although Griffiths et al. emphasize teleological explanations, in practice the success of a Bayesian model is taken as evidence for *both* teleological explanations and optimality claims.

Indeed, it is not always clear that the teleological explanation being offered is anything other than that human performance is near optimal. For example, consider the Bayesian reader model that we discussed in the target article (Bowers & Davis, 2012).

Jeffrey S. Bowers, School of Experimental Psychology, University of Bristol, Bristol, England; Colin J. Davis, Department of Psychology, Royal Holloway University of London, London, England.

We would like to thank Elizabeth Wonnacott for her helpful comments on a draft of this reply.

Correspondence concerning this article should be addressed to Jeffrey S. Bowers, School of Experimental Psychology, University of Bristol, 12a Priory Road, Clifton, Bristol BS8 1TU, England. E-mail: j.bowers@bristol.ac.uk

Norris (2006) wrote: "Both the general behavior of the model and the way the model predicts different patterns of results in different tasks follow entirely from the assumption that human readers approximate optimal Bayesian decision makers" (p. 327). This passage appears to highlight the importance of optimality per se, rather than any other teleological explanation. Indeed, the assumptions (priors, likelihood functions) do not appear to constitute a key theoretical claim. For instance, in order for the model to capture human performance in the lexical decision task, Norris assumed that there are "background" and "virtual nonwords" that are stored in memory, and the task of the participant is to decide whether a given input is closer to a word or a virtual nonword. A subsequent version of the model adopted another method for making lexical decisions, but in neither case were the likelihood functions presented as a central claim. Similarly, the Bayesian reader included word frequency as its prior, but again, this was not considered a key claim, as Norris noted that the prior could incorporate other factors such as age of acquisition if the data demanded it. What is constant across the iterations of the Bayesian reader model is the claim that word identification is near optimal. Or to give another example, to make their Bayesian model of motion perception better match human performance, Weiss, Simoncelli, and Adelson (2002) added a "gain control" function that mapped stimulus contrast into perceived contrast. The authors did not highlight this aspect of their model, and indeed, it does not appear to constitute a core part of their theory. Rather, the emphasis was placed on the claim that motion perception is near optimal, as reflected in the title of their article ("Motion illusions as optimal percepts").

Furthermore, even when theorists place teleological explanations at the forefront and relegate optimality claims to the background, it is not at all clear that a successful Bayesian model provides strong evidence for a given set of priors or likelihoods. That is, it is quite likely that another teleological explanation (another set of likelihoods and priors) will allow another Bayesian model to account for the data just as well (cf. Jones & Love, 2011). In other words, the success of a given model may be nothing more than a "just-so story." In the target article we gave numerous examples in which Bayesian theories were built post hoc around the data.

Griffiths et al. (2012) take issue with our claim that Bayesian models have more degrees of freedom than non-Bayesian models and describe our claim regarding the unfalsifiability of the Bayesian *framework* as misconceived. However, we never claimed that Bayesian models have more degrees of freedom compared to alternative approaches. In fact, we readily acknowledge the free parameters in all models, noting simply that "we have emphasized the flexibility of Bayesian models because it is often assumed (and claimed) that Bayesian models are more constrained than the non-Bayesian alternatives" (p. 397). We agree that when evaluating the Bayesian framework, the critical question is whether Bayesian models are productive or unproductive. Answers to this question can only be determined in the standard manner: namely, by comparing models and determining which of them provides a more parsimonious account of a wider range of data (as well as which models make novel predictions). As we detailed in Bowers and Davis (2012), Bayesian models are rarely compared to alternative theories (we explicitly claimed "rarely" rather than "never"), and when models are compared, the Bayesian models are often compared to "straw man" theories (e.g., Lewandowsky, Griffiths, & Kalish, 2009). The common failure to compare Bayesian to non-Bayesian models can easily lead to just-so stories in which performance is described as near optimal. For example, Wozny, Beierholm, and Shams (2010) argued that probability matching can be explained as near optimal on the assumption that participants are looking for predictable patterns. What they did not consider, however, is that the pattern can be explained by a suboptimal adaptive network approach that we discussed in some detail.

Theoretical Bayesians

Jones and Love (2011) challenged the *fundamentalist Bayesian* perspective, according to which theories are developed through "rational analysis" as described by Anderson (1990). That is, a theory is developed at Marr's computational level through a consideration of the problem and the environment, with little consideration of the underlying mechanisms (at the level of the brain) that support performance. This was clearly Anderson's approach, who wrote that the goal of rationalist theory "is to predict behavior from the structure of the environment rather than the structure of the mind" (Anderson, 1991, pp. 473–474), and that a rationalist theory "focuses us on what is outside the head rather than what is inside" (Anderson, 1990, p. 23). Nevertheless, in a response to Jones and Love, Chater et al. (2011) wrote: "Bayesian Fundamentalism is purely a construct of [Jones and Love's] imagination" (p. 194).

In Bowers and Davis (2012) we challenged a more modest interpretation of Bayesian theorizing, which we called the theoretical Bayesian approach. From this view, modelers are free to constrain their theories using behavioral data, biology, and the like. The key point is that theoretical Bayesians are making not only teleological claims (answering *why* questions) but also claims about how the mind and brain work at the algorithmic and implemental levels of description. Or as Chater, Oaksford, Hahn, and Heit (2010) wrote: "Bayesian methods . . . may bridge across each of Marr's levels of explanation" (p. 820). Nevertheless, according to Griffiths et al. (2012), there are few, if any, theoretical Bayesians either.

However, the claim that there are few, if any, theoretical Bayesians is hard to reconcile with many relevant statements in the literature. For instance, how are we to interpret the Körding and Wolpert (2006) quote at the start of Bowers and Davis (2012)? Again, one need only to look at the titles of many articles and books to understand where we got the impression that theorists often claim that the mind and brain actually perform (or approximate) Bayesian computations (e.g., Doya, Ishii, Pouget, & Rao, 2007). In addition, the frequent references to neuroscience provide clear evidence that researchers often take a theoretical Bayesian perspective (see Bowers & Davis, 2012).

Consider a recent article by Tenenbaum, Kemp, Griffiths, and Goodman (2011). These authors claimed that Bayesian models should be considered proposals about how to answer three central questions: (1) How does knowledge guide learning and inference? (2) What form does knowledge take? (3) How is knowledge acquired? That is, the authors appear to be focusing more on *how* rather than teleological *why* questions. The answers of Tenenbaum et al. to these questions appear to constitute algorithmic claims regarding representations and processes:

> Bayesian models typically combine richly structured, expressive knowledge representations (question 2) with powerful statistical inference engines (questions 1 and 3), arguing that only a synthesis of sophisticated approaches to both knowledge representation and inductive inference can account for human intelligence. (p. 1279)

Similarly, Tenenbaum et al. appear to take a theoretical Bayesian position when they write:

> Much ongoing work is devoted to pushing Bayesian models down through the algorithmic and implementation levels. The complexity of exact inference in large-scale models implies that these levels can at best approximate Bayesian computations, just as in any working Bayesian AI system. (p. 1284)

Given that Bayesian AI (artificial intelligence) systems typically make estimates of priors and likelihoods, multiply these probability functions, and multiply priors and likelihoods for at least some alternative hypotheses, it is not clear how to reconcile the above quote with Griffiths et al. (2012), who claim that there are few, if any, theoretical Bayesians. Perhaps we have misinterpreted Tenenbaum et al. (2011), among others, but we expect we are not the only ones.

Methodological Bayesians

The rapid progress in mathematics and computer science has allowed Bayesian analyses to be applied to a range of questions in psychology and neuroscience, but we question the value of these analyses if it is also assumed that the algorithms and mechanisms that mediate behavior are non-Bayesian. Putting aside the above concerns regarding post hoc theorizing, it is not clear how a detailed characterization of the optimal solution is an important advance over standard (non-Bayesian) adaptive solutions. Bayesian and adaptive solutions to a problem will inevitably be highly correlated (e.g., both Bayesian and non-Bayesian theories will take advantage of past learning—i.e., priors). The optimal solution will generally differ in detail from an adaptive solution (e.g., a solution derived from a neural network), but then, human performance generally differs in detail from optimal solutions. Griffiths et al. (2012) suggest that observing a correspondence between an optimal solution and human behavior "suggests that we should begin to explore approximate algorithms that can find decent solutions to these problems in reasonable time" (p. 416). But we would argue that this type of exploration is already being undertaken by computational modelers of cognition, without necessitating a prior Bayesian analysis.

Of course, if there were cases in which a Bayesian model identified a counterintuitive solution to a problem—or better, made a novel prediction that current non-Bayesian models miss—then the model would be making an important contribution. But this is rarely the case. Or if Bayesian models were needed in order to make teleological explanations, we would agree that the approach is essential. But it seems to us that stronger teleological conclusions can be made with traditional non-Bayesian process models. That is, the teleological explanations derived from processing models consider the role that algorithms play in explaining why performance is as it is.

Bayesian Neuroscience

Given that Griffiths et al. (2012) reject the view that the brain is actually performing (or approximating) Bayesian computations, it is surprising that they take issue with our claim that there is little or no evidence from neuroscience in support of Bayesian theories. They consider our conclusion "remarkably assertive," but our position should not be considered so extreme given that two of the key advocates of the Bayesian coding hypothesis wrote: "The neurophysiological data on the [Bayesian coding] hypothesis, however, is almost non-existent" (Knill & Pouget, 2004, p. 712). Accordingly, our strong claims seem well justified when considering the data collected up to that point. To support our conclusion, we looked at more recent evidence and found little or no reason to update this view. Our conclusion regarding the neuroscience is not so different from that of Vilares and Körding (2011), who wrote in a recent review of Bayesian models:

> Finally, we want to point out that there is currently some disconnection between Bayesian theories and experimental data about the nervous system. While there are many theoretical proposals of how the nervous system might represent uncertainty, there is not much experimental support for any of them. (p. 35)

The main criticism of Griffiths et al. (2012) about our analysis is that we assumed that the Bayesian coding hypothesis depends on the intrinsic noisiness of neurons, which they take as a significant misreading of Ma, Beck, Latham, and Pouget (2006). However, as Ma and Pouget (2009) wrote: "An important aspect of population codes is the stochastic nature of neuronal responses" (p. 749), and this internal noise is thought to provide ambiguities that are addressed through Bayesian solutions (e.g., Orbán & Wolpert, 2011). Furthermore, contrary to Griffiths et al., we did not suggest that the Bayesian coding hypothesis rests on the idea that neurons are stochastic devices. Rather, we wrote: "Of course, even if single neurons are more reliable than assumed by Ma et al. (2006), this does not rule out the claim that collections of neurons compute in a Bayesian-like fashion" (Bowers and Davis, 2012, p. 404).

There is one way in which a Bayesian theory of neural computation might have stronger support: that is, if Bayesian theories are characterized so broadly as to encompass almost any theory of neural signaling. For example, Knill and Pouget (2004) also wrote:

> This is the basic premise on which Bayesian theories of cortical processing will succeed or fail—that the brain represents information probabilistically, by coding and computing with probability density functions or approximations to probability density functions. . . . The opposing view is that neural representations are deterministic and discrete. . . . (p 713)

This is the same point that Griffiths et al. (2012) make when they say the critical question for Bayesian theory of neuroscience is whether neurons code single values or probability distributions. If we understand this characterization of Bayesian neural processing, all that is required is that a given input produces a pattern of activation across a set of neurons, and this pattern (rather than the discrete output of single neurons) influences higher levels of processing before a decision is made. We would agree that there is long-standing evidence for this supposition, both behavioral and neurobiological. Indeed, the claim that cognitive and neural representations are deterministic and discrete (the "non-Bayesian" approach) is adopted by no one, as far as we are aware. The view that graded signals pass on their signal to higher levels of processing is a common claim in most theories in psychology that goes by the name of "cascaded processing."

Conclusions

What is clear from the recent debate started by Jones and Love (2011) and continued by us is that there is a good deal of confusion about what theoretical claims are being advanced by Bayesian modelers. Jones and Love thought that many Bayesian theorists were claiming that much about the mind could be understood by studying what is outside the head (i.e., the environment and the task) rather than what is inside the head, but this Bayesian fundamentalist position was strongly rejected by leading Bayesian theorists. We thought that claims regarding optimality were novel and central to the Bayesian approach, and that many Bayesian theorists were claiming that cognition and behavior were supported by Bayesian-like algorithms, but this theoretical Bayesian perspective has also been rejected by Griffiths et al. (2012). There seems to be agreement that there are Bayesian theorists who adopt what we refer to as a methodological Bayesian approach, although there is disagreement as to the promise of this approach (of which we are doubtful). We also suspect that a clear statement of the methodological Bayesian position would have generated far less excitement than a theoretical Bayesian perspective that researchers may have been (mis)understood to be embracing. Given the recent disagreements regarding the fundamentalist and theoretical Bayesian perspectives (and whether these views even exist), there is a pressing need to clarify what Bayesian theories of cognition are all about.

References

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

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences, 14,* 471–485. doi:10.1017/S0140525X00070801

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

Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1,* 811–823. doi:10.1002/wcs.79

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (Eds.). (2007). *Bayesian brain: Probabilistic approaches to neural coding.* Cambridge, MA: MIT Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. doi:10.1038/415429a

Faisal, A. A., & Wolpert, D. M. (2009). Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception–action task. *Journal of Neurophysiology, 101,* 1901–1912. doi:10.1152/jn.90974.2008

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

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin, 138,* 415–422. doi:10.1037/a0026884

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

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27,* 712–719. doi:10.1016/j.tins.2004.10.007

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10,* 319–326. doi:10.1016/j.tics.2006.05.003

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

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., & Pouget, A. (2009). Population codes: Theoretic aspects. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (Vol. 7, pp. 749–755). Oxford, England: Academic Press.

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

Orbán, G., & Wolpert, D. M. (2011). Representations of uncertainty in sensorimotor control. *Current Opinion in Neurobiology, 21,* 629–635. doi:10.1016/j.conb.2011.05.026

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331,* 1279–1285. doi:10.1126/science.1192788

Vilares, I., & Körding, K. P. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences, 1224,* 22–39. doi:10.1111/j.1749-6632.2011.05965.x

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5,* 598–604. doi:10.1038/nn0602–858

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology, 6,* e1000871. doi:10.1371/journal.pcbi.1000871

Received January 8, 2012
Revision received January 18, 2012
Accepted January 19, 2012 ■