Bayesian Just-So Stories in Psychology and Neuroscience

Key Words: Bayes; Bayesian; optimal; heuristics; just-so stories

Jeffrey S. Bowers
School of Experimental Psychology
University of Bristol
j.bowers@bris.ac.uk

Colin J. Davis
Department of Psychology
Royal Holloway
C.Davis@rhul.ac.uk

Abstract

According to Bayesian theories in psychology and neuroscience, minds and brains are (near) optimal in solving a wide range of tasks. We challenge this view and argue that more traditional, non-Bayesian approaches are more promising. We make three main arguments. First, we show that the empirical evidence for Bayesian theories in psychology is weak at best. This weakness relates to the many arbitrary ways that priors, likelihoods, and utility functions can be altered in order to account for the data that are obtained, making the models unfalsifiable. It further relates to the fact that Bayesian theories are rarely better at predicting data compared to alternative (and simpler) non-Bayesian theories. Second, we show that the empirical evidence for Bayesian theories in neuroscience is weaker still. There are impressive mathematical analyses showing how populations of neurons could compute in a Bayesian manner but little or no evidence that they do. Third, we challenge the general scientific approach that characterizes Bayesian theorizing in cognitive science. A common premise is that theories in psychology should largely be constrained by a rational analysis of what the mind ought to do in order to perform optimally. We question this claim and argue that many of the important constraints come from biological, evolutionary, and processing (algorithmic) considerations that have no adaptive relevance to the problem per se. In our view, these factors have contributed to the development of many Bayesian "just-so" stories in psychology and neuroscience; that is, mathematical analyses of cognition that can be used to explain almost any behavior as optimal.

In recent years there has been an explosion of research directed at a surprising claim: namely, that minds and brains are (near) optimal in solving a wide range of tasks. This hypothesis is most strongly associated with Bayesian theories in psychology and neuroscience that emphasize the statistical problems confronting all organisms. That is, cognitive, motor and perceptual systems are confronted with noisy and ambiguous inputs—e.g., a 3D world is projected on a 2D retina—and, from this view, these systems are designed to carry out or approximate Bayesian statistics in order to make optimal decisions given the degraded inputs. Typical conclusions include the following:

> *…it seems increasingly plausible that…in core domains, human cognition approaches an optimal level of performance.* (Chater, Tenenbaum & Yulle, 2006, p. 289)

> *These studies… have shown that human perception is close to the Bayesian optimal suggesting the Bayesian process may be a fundamental element of sensory processing.* (Körding & Wolpert, 2006, p. 321)

> *One striking observation from this work is the myriad ways in which human observers behave as optimal Bayesian observers. This observation… has fundamental implications for neuroscience, particularly in how we conceive of neural computations and the nature of neural representations of perceptual and motor variables*. (Knill & Pouget, 2004, p. 712)

> *Our results suggest that everyday cognitive judgments follow the same optimal statistical principles as perception and memory* (Griffiths, & Tenenbaum, 2006, p. 767).

These conclusions are exciting, not only because they are counterintuitive (who would have thought we are optimal?), but also because they appear to constitute novel claims about mind and brain. In the standard view, cognitive, perceptual, and motor systems are generally good at solving important tasks, but the limitations of the systems were always salient. For example, various heuristics are often thought to support high-level reasoning and decision making. These heuristics are adaptive under many conditions, but not optimal (in fact, how far from optimal is a matter of some dispute; e.g., Gigerenzer & Brighton, 2009; Gigerenzer, Todd, & the ABC Research Group, 1999; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1996). In a similar way, perception is often characterized as a "bag of tricks" (Ramachandran, 1990). That is, the perceptual systems rely on heuristics that generally work well enough but in no way approximate Bayesian solutions. More generally, it is often assumed that evolution produces systems that *satisfice* (Simon, 1956), or *meliorize (*Dawkins, 1982). That is, selective adaptation produces "good enough" solutions, or "better than alternative" solutions but not optimal solutions. The Bayesian approach, by contrast, appears to claim that evolution has endowed us with brains that are exquisitely good at learning and exploiting the statistics of the environment, such that performance is close to optimal.

In this article we challenge the Bayesian approach to studying the mind and brain and suggest that more traditional, non-Bayesian approaches provide a more promising way to proceed. We organize our argument as follows. In Part 1 we introduce Bayesian statistics and summarize three different ways in which these methods have influenced theories in psychology and neuroscience. These different approaches to Bayesian theorizing make quite different claims regarding how the mind works. In Part 2, we highlight how there are too many arbitrary ways that priors, likelihoods, utility functions, etc. can be altered in a Bayesian model in order to account for the data that are obtained. That is, Bayesian models are difficult to falsify. Our concern is not just hypothetical concern; we describe a number of Bayesian models developed in a variety of domains that were built post-hoc in order to account for the data. In Part 3 we show how the predictions of Bayesian theories are rarely compared to alternative non-Bayesian accounts that assume that humans are reasonably good at solving problems (e.g., heuristic or adaptive theories of mind). This is problematic given that the predictions derived from optimizing (Bayesian) and adaptive (non-Bayesian) theories will necessarily be similar. We review a number of cases in a variety of domains in which data taken to support the Bayesian theories are equally consistent with non-Bayesian accounts.

Next, in Part 4, we consider the claim that collections of neurons perform Bayesian computations. We argue that the data in support of this claim are weaker still. There are impressive mathematical analyses showing how populations of neurons should compute in order to optimize inferences given certain types of noise (variability) in neural responding, but little or no evidence exists that neurons actually behave in this way. Finally, in Part 5, we challenge the general scientific approach that characterizes most Bayesian theorizing in cognitive science. A key premise that underlies most Bayesian modelling is that the mind can be studied by focusing on the environment and the task at hand, with little consideration of what goes on inside the head. That is, the most important constraints for theories of mind can be discovered through a rational consideration of what a mind ought to do in order to perform optimally. We question this claim and argue that many of the important constraints come from processing (algorithmic), biological, and evolutionary considerations that have no adaptive relevance to the problem per se. Not only does this “rational” approach to cognition lead to under-constrained theories, it dissociates theories of cognition from a wide range of empirical findings in psychology and neuroscience.

In our view, the flexibility of Bayesian models, coupled with the common failure to contrast Bayesian and non-Bayesian accounts of performance, has led to a collection of Bayesian “just-so” theories in psychology and neuroscience: Sophisticated statistical analyses that can be used to explain almost any behavior as (near) optimal. If the data had turned out otherwise, we contend, a different Bayesian theory would have been carried out to justify the same conclusion, that is, that the mind/brain supports near optimal performance.

Part 1: What is a Bayesian theory in psychology (and neuroscience)?

At the most general level, Bayesian theories in cognitive psychology and neuroscience assume that the mind and brain perform Bayesian statistics, or something functionally similar in a given context. The premise is that cognitive and perceptual systems need to make decisions about unique events on the basis of noisy and ambiguous information. Bayesian statistics are the optimal method for estimating probabilities of unique events, and the mind/brain is assumed to apply or approximate this method in order to make optimal (or near optimal) decisions.

Most Bayesian theories are developed at a *computational* rather than an *algorithmic* level of description (Marr, 1982). That is, Bayesian theories describe the goal of a computation, why it is appropriate, and the logic of the strategy, but not the mental representations and processes that are employed in solving a task. Bayesian theories in psychology typically adopt the “rational analysis” methodology described by Anderson (1991). That is, the focus is on understanding the nature of the environment (e.g., what information is available to an organism) and the nature of the task being performed. Together, these factors, when combined with Bayesian probability theory, determine what an optimal solution should look like. Critically, this solution is thought to provide important constraints on theories of mind/brain.

To avoid any confusion, it is important to note that our criticism of the Bayesian approach has nothing to do with Bayesian statistics or Bayesian decision theory per se. That is, we do not take issue with the claim that Bayesian methods provide optimal methods for determining probabilities of specific events. Nor do we dispute the promise of Bayesian analysis methods in the evaluation of cognitive models (e.g., Lee, 2008; Rouder & Lu, 2005; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). What we do question, however, is the relevance of this statistical approach to theories of mind and brain.

Bayesian probability

Before considering the virtues of Bayesian theories in psychology and neuroscience, it is perhaps worth reviewing the basics of Bayesian statistics and their use in optimal decision making. Although specific Bayesian methods can be quite complicated, it is important to have a general understanding of Bayesian statistics, given the claim that the mind and brain in some way implement or approximate these methods.

Bayes's theorem specifies the optimal way of combining new information with old information. More specifically, if we have some hypothesis about the world, possibly based on prior information, Bayes's rule tells us to how to re-evaluate the probability of this hypothesis in the light of new evidence. The rule itself is quite straightforward, and may be written in the form:

$$P\ (H|E) = P\ (H) \text{ x } P(E|H)\ /\ P(E) \qquad (1)$$

Here, *H* is the hypothesis under investigation and *E* is the new evidence. The lefthand side of the equation, *P(H|E)*, is called the posterior probability. This represents the probability that the hypothesis is true given the new evidence (the symbol | means "given"). The first term on the righthand side of the equation, *P(H)*, is called the prior probability. This represents the prior probability that the hypothesis is true before the new evidence is taken into account. The second term on the right-hand side of the equation, *P(E|H)*, is called the likelihood function. This represents how likely it is that the new evidence would have been obtained given that the hypothesis *H* is true. Finally, the denominator *P(E)* represents the probability of the new evidence. Because this term is constant with respect to the hypothesis *H* it can be treated as a scaling factor and we can write the following:

$$\text{Posterior probability} \propto \text{Likelihood function x Prior probability}$$

where the symbol $\propto$ is read "is proportional to".

To illustrate the components of this formula and how Bayes's rule is used to estimate the posterior probability, consider the following concrete example. Imagine that the hypothesis under consideration is that a 30 year old man has lung cancer, and the new evidence is that he has a cough. Accordingly, the posterior probability we want to calculate is the probability that the man has cancer given that he has a cough, or *P(cancer|cough)*. Adapting equation (1) to the current situation, we have:

$$P\,(cancer|cough) = P\,(cancer) \text{ x } P(cough|cancer) \,/\, P(cough) \qquad (2)$$

For this example, let us assume we have exact values for the priors and likelihoods. Namely, the prior probability of a 30 year old man in the relevant population having lung cancer is .005, the probability of coughing when you have cancer is .8 (likelihood), and the overall probability of coughing is .2. Note, the symptom of coughing can have many different causes, most of which are not related to lung cancer, and the probability of 0.2 ignores these causes and simply represents how likely it is that someone will have a cough for whatever reason. Plugging these numbers into Equation (2) leads to the following:

$$P\ (cancer|cough) = .005 \text{ x } 0.8 / 0.2 = 0.02$$

The point to note here is that the estimated probability of cancer has only gone up from .005 to .02, which makes intuitive sense given that coughing is not a strong diagnostic test for cancer.

However, the resulting posterior probabilities are not always so intuitive. Imagine that instead of relying on coughing, the relevant new data is a patient's positive result on a blood test that has a much higher diagnostic accuracy. To evaluate this evidence it is helpful to decompose *P(E)* as follows:

$$P(E) = P\ (H) \text{ x } P(E|H) + P\ (\sim H) \text{ x } P(E|\sim H) \qquad (3)$$

where ~ means "not true". That is, there are two possible explanations of the positive result. The first possibility is that the result is a correct detection. The probability of this is $P(cancer)$ x $P(positive\ test\ |cancer)$, where $P(positive\ test\ |cancer)$ represents the hit rate (or sensitivity) of the test. Suppose that this hit rate is .995. The second possibility is that the result is a false alarm. The probability of this is $P(\sim cancer)$ x $P(positive\ test\ |\sim cancer)$, where $P(positive\ test\ |\sim cancer)$ represents the false alarm rate of the test, i.e., the probability of the test's being positive when you do not have cancer. Suppose that this false alarm rate is 0.01. We can infer from these rates that the overall probability of a positive test result is *P(cancer)* x *Hit rate* + *P(~cancer)* x *False alarm rate* = .005 x .995 + .995 x .01 = 0.015.

Plugging these numbers into Bayes's rule leads to the following:

$$P\ (cancer|positive\ test) = .005 \text{ x } 0.995 \,/\, (.005 \text{ x } .995 + .995 \text{ x } .01) = 0.33$$

That is, even though the test has 99.5% detection accuracy when cancer is present, and has 99% accuracy when it is not, the probability that the patient has cancer is only 33%. Why is this? It reflects the fact that posterior probabilities are sensitive to both likelihoods and prior probabilities. The reason that the estimate of cancer given the blood test is higher than the estimate of cancer given a cough is that the likelihood has changed. And the reason that *P* (*cancer|positive test*) is much lower than 99% is that the prior probability of cancer is so low. To see the influence of the prior, consider the same test result when the prior probability of cancer is much higher, say 1 in 3 (e.g., imagine that the person being tested is a 90 year old lifetime smoker). Then the computed probability is

$$P\ (cancer|positive\ test) = .333 \text{ x } 0.995 \,/\, (.333 \text{ x } .995 + .667 \text{ x } .01) = 0.98$$

The important insight here is that the very same test producing the same result is associated with a very different posterior probability because the prior probability has changed. The necessity of considering prior probabilities (base rates) is one that most people find quite counter-intuitive. Indeed, doctors tend to perform very poorly at determining the correct posterior probability in real-world problems like the above. For example, Steurer, Fischer, Bachmann, Koller and ter Riet (2002) found that only 22% of general practitioners were able to use information about base rate and test accuracy to correctly estimate the (low) probability of a patient having a disease following a positive test result. Indeed, the authors suggested that their findings might overestimate the average performance of general practitioners, as their participants were recruited from doctors attending courses on evidence-based medicine. Similarly, Eddy (1982) reported that 95 out of 100 doctors estimated the posterior probability to be approximately equal to the likelihood *P*(*E*|*H*), apparently ignoring the prior probability *P*(*H*). This assumption would be equivalent to reporting a 99.5% probability instead of a 33% probability (see Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007 for discussion of the conditions that facilitate Bayesian reasoning).

In the Bayesian models discussed in this article it is often the case that there are multiple hypotheses under evaluation. For example, in a model of word identification, there may be separate hypotheses corresponding to each possible word. We can then make use of the Law of Total Probability, which says that if the set of hypotheses H_1, H_2, …, H_n is exhaustive, and each of these hypotheses is mutually exclusive, then:

$$P(E) = \sum_{i=1}^{n} P(H_i) \text{ x } P(E|H_i) \tag{4}$$

That is, the probability of the evidence can be found by summing the joint probabilities over all hypotheses (where the joint probability is the prior probability times the likelihood), a process referred to as marginalization. In practice, it may not be possible to guarantee the exclusivity and exhaustivity of a set of hypotheses. However, for many practical purposes this does not matter, as we are typically interested in the relative probability of different hypotheses (i.e., is this new evidence best explained by H_1 or H_2?), meaning that the absolute value of the scaling factor $P(E)$, which is independent of any specific hypothesis H_i, is not critical.

The value of using Bayes's rule to make rational judgments about probability is clear in the case of clinical examples like cancer diagnosis. The strong claim made by proponents of Bayesian inference models of cognition and perception is that humans make use of the same rule (or close approximations to it) when perceiving their environment, making decisions, and performing actions. For example, while walking across campus, suppose you see someone who looks like your friend John. In the Bayesian formulation, "looks like John" might correspond to a specific likelihood, e.g., P(E | John) = 0.8. Is it actually John? A rational answer must take into account your prior knowledge about John's whereabouts. If John is your colleague, and you often see him on campus, the chances are quite high that the person you have seen is in fact John. If John is your next-door neighbor, and you've never seen him on campus, the prior probability is lower, and you should be less confident that the person you have seen is John. If your friend John died 20 years ago, the rational prior probability is 0, and thus you should be perfectly confident that the person you saw is not John, however much he may resemble him (i.e., whatever the likelihood). Of course, the idea that previous knowledge influences perception is not novel. For example, there are abundant examples of top-down influences on perception. What is unique about the Bayesian hypothesis is the claim about how prior knowledge is combined with evidence from the world, i.e., the claim that humans combine these sources of information in an optimal (or near optimal) way, following Bayes's rule.

There are a number of additional complications that should briefly be mentioned. First, unlike the simple examples above, the priors and likelihoods in Bayesian theories of the mind generally take the form of probability distributions rather than unique estimates, given noise in the estimates. For example, the estimate of the likelihood of a given hypothesis might be a distribution centered around a probability of 0.8, rather than consisting of a single point estimate of 0.8. This is illustrated in Figure 1, in which there is uncertainty associated with both the likelihood and the prior, resulting in a distribution for the posterior probability; as can be seen, the posterior distribution is pulled in the direction of the prior (or, to put it another way, the prior is updated to the posterior by being shifted in the direction of the data). Thus, Bayesian methods enable decision makers to go beyond point estimates and take into account the uncertainty associated with a test in order to make optimal decisions. As we will see below, however, the practice of estimating priors and likelihoods is fraught with difficulties.

Second, posterior probabilities do not always determine what the optimal decision is. Imagine that the posterior probability of cancer is best characterized as a probability distribution centered around 10%. What is one to make of this fact? It obviously depends on other factors (e.g., life expectancy associated with the cancer, the risks and benefits associated with treatment, costs, etc.), as well as the shape of probability distribution (e.g., how tightly centred around 10% is the probability distribution? Is the distribution skewed?). Accordingly, in Bayesian optimal decision theory, the posterior probability is combined with a utility function, so that all the relevant variables (including the probability of cancer) can be considered when making a decision. The optimal decision is the one that maximize utility—or equivalently, minimizes loss. However, this decision will depend on which variables affect the utility function and how the resulting utility function relates to the probability of cancer. For example, surgery may be the best option if the probability of cancer is very high, whereas some other treatment may maximize utility if the probability is moderate. If the best estimate of the probability of cancer is .02, the best decision is to have no treatment. As noted below, the estimation of utility functions in psychological models is not straightforward, as both the variables that affect utility and the shape of the function relating these variables to utility tend to be unknown.

A possible confusion regarding "optimality".

Undoubtedly one of the reasons that Bayesian theories have gathered such attention in psychology and neuroscience is that claims of optimality are counter-intuitive. Indeed, such claims might at first sound absurd. If humans are optimal, why are we unable to outrun a cheetah, outswim a dolphin, or fly? Indeed, why don't we run, swim and fly at the limits set by physics? What needs to be emphasized is that optimality means something quite specific in the Bayesian context. The core claim is that the mind makes or approximates optimal decisions given noisy data. The noise that the brain has to deal with is not only the product of physical world itself but also suboptimal human design (e.g., it is suboptimal that photoreceptors are located at the back of the retina, and that neurons fire differently to the same stimulus on different occasions, etc.). But bad design is not problematic for Bayesian theories claiming that human cognition approaches optimality. In principle, there is no limit on how badly designed systems are, except in one respect. The decision stage that interprets noisy, suboptimal inputs is hypothesized to be near optimal.

To sum up, Bayesian statistics provide a method of computing the posterior probability of a hypothesis, which provides the best way to update a prior belief given new evidence. Bayesian decision theory defines how our beliefs should be combined with our objectives to make optimal decisions given noisy data.

Three types of Bayesian theorizing in psychology and neuroscience

Three different approaches to Bayesian theorizing in psychology and neuroscience can be distinguished; we refer to these as the Extreme, Methodological, and Theoretical approaches. The difference concerns how Bayesian theories developed at a computational level of description relate to the algorithmic level. On the Extreme view, psychology should only concern itself with developing theories at a computational level. Questions of how the mind actually computes are thought to be intractable and irrelevant. For instance, Movellan and Nelson (2001) write:

> *Endless debates about undecidable structural issues (modularity vs. interactivity, serial vs. parallel processing, iconic vs. propositional representations, symbolic vs. connectionist models) may be put aside in favor of a rigorous understanding of the problems solved by organisms in their natural environments.* (pp. 690-691)

We do not have much to say about this approach, as it dismisses the questions that are of interest to most cognitive psychologists. This approach might make good sense in computer science or robot labs, but not, in our view, cognitive science. In any case, such extreme views are rare.

More common is the Methodological Bayesian approach. Methodological Bayesians are not committed to any specific theory of mind, at any level of description. Rather, they use Bayesian models as tools: The models provide a measure of optimal behavior that serve as a benchmark for actual performance. From this perspective, the striking result is how often human performance is near optimal, and this is considered useful for constraining a theory (whatever algorithm the mind uses, it must support behavior that approximates optimal performance). But this approach is in no way committed to the claim that the mind/brain computes in a Bayesian-like way at the algorithmic level. For example, Geisler and Ringach (2009) write:

> *There are a number of perceptual and motor tasks where humans parallel the performance predicted by ideal Bayesian decision theory; however, it is unknown whether humans accomplish this with simple heuristics or by actually implementing the machinery of Bayesian inference.* (p. 3)

That is, the Methodological approach is consistent with qualitatively different algorithmic descriptions of mind. On the one hand, the mind might be Bayesian-like and compute products of priors and likelihoods for all of the possible hypotheses (as in Equation 1). On the other hand, the mind might be distinctly non-Bayesian,and carry out much simpler computations, considering only a small subset of the available evidence to reach a near optimal decision.

Theoretical Bayesians agree that Bayesian models developed at the computational level constitute a useful method for constraining theories at the algorithmic level (consistent with the Methodological Bayesian approach), but in addition, claim that the mind carries out or approximates Bayesian computations at the algorithmic level, in some unspecified way. For example, when describing the near optimal performance of participants in making predictions of uncertain events, Griffiths and Tenenbaum (2006) write: “These results are inconsistent with claims that cognitive judgments are based on non-Bayesian heuristics." (p. 770). Indeed, Bayesian models developed at a computational level are thought to give insight into how neurons compute. For instance, Kording and Wolpert (2006, p. 322) write:

> *…over a wide range of phenomena people exhibit approximately Bayes-optimal behaviour. This makes it likely that the algorithm implemented by the [central nervous system] may actually support mechanisms for these kinds of Bayesian computations*. (p. 322)

When Chater, Oaksford, Hahn, and Heit (2010) write "Bayesian methods… may bridge across each of Marr's levels of explanation" (p. 820), we take it that they are adopting a Theoretical Bayesian perspective in which computational, algorithmic, and implementational descriptions of the mind may all be Bayesian.

Again, Theoretical Bayesians are not committed to specific claims regarding how the mind/brain realizes Bayesian computations at the algorithmic and implementational levels. But in some way a Bayesian algorithm 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 (the denominator in Equation 1). It is a commitment to the above four claims that makes a theory Bayesian (as opposed to simply adaptive), and a novel claim about how the mind/brain works.[1]

[1] Our distinction between Methodological and Theoretical Bayesian theories is similar to the distinction that Brighton and Gigerenzer (2008) make between "broad" and "narrow" senses of the probabilistic mind. Jones and Love (2011) categorize Bayesian theories somewhat differently, subdividing theories into what they call Bayesian Fundamentalism and Bayesian Enlightenment. The Bayesian Fundamentalism category is similar to what we call the Extreme Bayesian approach, whereas Bayesian Enlightenment could encompass both the Methodological and Theoretical approaches (cf., Bowers & Davis, 2011).

In principle, the contrast between Methodological and Theoretical Bayesian theories is straightforward. However, two factors act to blur this distinction in practice. First, theoretical Bayesians are not committed to the claim that the algorithms of the mind are perfectly Bayesian. Many problems are computationally intractable when framed in pure Bayesian terms, and thus it is necessary to rely on various approximations to the prior and likelihood distributions, as well as approximations to the hypothesis space. Accordingly, the challenge from the Theoretical perspective is to develop psychologically plausible algorithms that provide good estimates of priors, etc. while relying on limited resources. Although these approximations are sometimes called heuristics (e.g., Sanborn, Griffiths, & Navarro, 2010), they are qualitatively different from the heuristics as envisaged by, for example, the "bias and heuristics" paradigm (e.g., Kahneman & Tversky, 1982) or the "ecological rationality" approach (Gigerenzer, Todd, and the ABC Research Group, 1999), according to which the mind relies on simple (sometime non-probabilistic) shortcuts in lieu of priors, likelihoods, and optimal methods of combining probability distributions. Again, the output of a good heuristic model might approximate the decisions of a Bayesian model in a given context, but it will not approximate the underlying processes.

Second, researchers are often inconsistent or unclear with regards to their position. For instance, in various passages in various publications, Oaksford and Chater (2003, 2007) argue that the mind is most likely a probabilistic device, cite findings from neuroscience that suggest that the collections of neurons can carry out probabilistic calculations (see Part 4 below), and argue that human everyday reasoning is too flexible for heuristic solutions:

> *…the onus is on the advocates of an across-the-board view that human reasoning is no more than a collection of reasoning heuristics to show how the flexibility of human reasoning is possible. (Oaksford & Chater, 2007, p. 278)*

At other times, they claim that Bayesian and heuristic theories are quite consistent with each other. Indeed, in places, they have endorsed algorithmic models that are non-Bayesian and constrained by satisficing rather than optimality criteria:

> *We suspect that, in general, the probabilistic problems faced by the cognitive system are simply too complex to be solved directly, by probabilistic calculation. Instead, we suspect that the cognitive system has developed relatively computationally "cheap" methods for reaching solutions that are "good enough" probabilistic solutions to be acceptable.* (Oaksford & Chater, 2007, p. 83)

The failure to clearly distinguish between the Methodological and Theoretical approaches is not inconsequential, for at least two reasons. First, it makes it difficult to appreciate what theoretical issues are at stake. Indeed, it is not always clear whether any theoretical issues are salient. For example, according to Chater (2000): "…rational methods can be views as compatible with the 'ecological' view of rationality outlined in Gigerenzer" (p. 746), whereas Brighton and Gigerenzer (2008) strenuously disagree. It is hard to have a useful debate if the nature of the disagreement is unclear. Second, it makes it difficult to evaluate current claims given that the Methodological and Theoretical Bayesian approaches need to be tested in different ways. If researchers are claiming that Bayesian models constitute an important methodological tool for formulating and evaluating theories, then it is only necessary to evaluate how this method provides important constraints for theorizing above and beyond previous methods. However, if the claim is that perception, cognition, and behavior are supported by Bayesian-like algorithms, then it is necessary to show that Bayesian theories are more successful than alternative theories. Below we review the literature in the context of the Methodological and Theoretical Bayesian distinction, and challenge the contribution of both approaches.

Part 2: Bayesian theories are flexible enough to account for any pattern of results

Bayesian theories of mind and brain have been developed in a wide range of domains, from high level cognition to low level perception and motor control. It is our contention that the evidence presented in support of Bayesian theories in these various domains is much overstated, for two reasons. The first reason, considered in this section, is that there are too many arbitrary ways that priors, likelihoods, utility functions, etc. can be altered in a Bayesian theory post-hoc. This flexibility allows these models to account for almost any pattern of results.

Speed perception

Weiss, Simonceilli, and Adelson (2002) developed a Bayesian model of motion perception that accounts for an illusion of speed: Objects appear to move more slowly under reduced or low contrast conditions. In order to accommodate these findings within a Bayesian framework, Weiss et al. assumed that bottom-up evidence is degraded under poor viewing conditions, and accordingly, the likelihood function is associated with greater uncertainty. The second assumption is that objects tend to move slowly in the world, and accordingly, the prior is biased towards slow movement. Together, these two assumptions can explain why objects appear to move more slowly under poor viewing conditions (because the prior plays a larger role in the computation of the posterior probability). Weiss et al. relate this analysis to a variety of findings, including the tendency of automobile drivers to speed up in foggy conditions.

Clearly, the prior is playing a key role in the model's ability to account for the illusion. The question, then, is whether this prior is motivated independently of the phenomenon the model is trying to explain. The answer is no, as articulated by Weiss et al. (2002) themselves:

> *We formalized this preference for slow speeds using a prior probability distribution … we have no direct evidence (either from first principles or from empirical measurements) that this assumption is correct. We will show, however, that it is sufficient to account qualitatively for much of the perceptual data. (p. 599)*

Furthermore, in order to account for perceptual data in a quantitative rather than qualitative manner, Weiss et al. (2002) introduced a nonlinear 'gain control' function that mapped stimulus contrast into perceived contrast. They provided no justification for this new likelihood function other than that it improved the fit of the model.

It should be noted that Weiss et al. (2002) assumed that the same motion prior is used when estimating the speed of computer generated gratings as when viewing natural objects from a moving car. But it would at least be possible that an optimal Bayesian system might conditionalize speed priors to different contexts, such that the speed prior of an oncoming car is not the same as for a leaf blowing in the wind. Indeed, many Bayesian models capitalize on different priors for different categorizes of objects in order to make their predictions (e.g., Hemmer & Steyvers, 2009), and Seydell, Knill and Trommershäuser (2010) highlight how quickly priors can change through experience. Assuming that all things move slowly, as in Weiss et al.'s prior, seems decidedly suboptimal.

More problematically, the illusion that objects appear to move slowly under poor illumination conditions does not occur under all conditions. For example, Thompson et al. (2006) found that although perceived speed is reduced for slow rates of movement, it is often overestimated for faster speeds.   Similar increases in perceived speed at low contrast have been reported by Hammett, Champion, Thompson, and Morland (2007), who note that this poses a challenge for Weiss et al.'s (2002) Bayesian theory. In response to these findings, Stocker and Simoncelli (2006) note that their Bayesian theory of speed perception could account for this phenomenon as well:

> *… if our data were to show increases in perceived speed for low-contrast high-speed stimuli, the Bayesian model described here would be able to fit these behaviors with a prior that increases at high speeds. (p. 583)*

But again, they do not provide any motivation for this new prior, other than to account for the data. In summary, there is little evidence for Weiss et al.'s (2002) assertion that humans perceive motion in the manner of ideal observers (for recent evidence regarding the "remarkable inefficiency" of humans' motion perception, see Gold, Tadin, Cook, & Blake, 2008).

Word identification

Norris and colleagues (Norris, 2006, 2009; Norris & Kinoshita, 2008; Norris, Kinoshita, & van Casteren, 2010; Norris & McQueen, 2008) have published a series of theoretical and empirical articles taken to support a Bayesian account of visual (and spoken) word identification. The models assume that readers and listeners are Bayesian decision makers, making optimal decisions on the basis of the available (noisy) evidence and their knowledge of the prior probabilities of words. Indeed, compared to alternative theories, the Bayesian reader is claimed to provide a better and more principled account of a range of phenomena, including the impact of word frequency on response times to identify words (Norris, 2006), the impact of form similarity on word identification in different tasks (Norris et al., 2010), and various priming phenomena (Norris & Kinoshita, 2008). However, the choices of priors, likelihoods, and decision rules are critical to the model's performance, and the specific choices of these parameters that have been adopted have been driven by empirical, rather than rational considerations.

In the initial formulation of the Bayesian Reader model, the prior probability for each word was based on its frequency in the language, enabling the model to provide a good account of word frequency effects. However, although word frequency is a large contributor to the time it takes to identify a word, it is not the only contributor. For example, another variable that influences word identification is age of acquisition (AoA; e.g, Brysbaert, Lange, & Van Wijnendaele, 2000; Morrison & Ellis, 1995; Stadthagen-Gonzalez, Bowers, & Damian, 2004). Norris (2006) argues that such variables can be incorporated into the priors, noting that "the Bayesian account is neutral as to the exact mechanism responsible for calculating prior probabilities" (p. 347); "frequency is not the only factor that can influence the prior probability of a word" (p. 331); and that factors such as cumulative frequency, AoA, and context "can be thought of as simply influencing the psychological estimate of a word's prior probability" (p. 334). That is, the Bayesian model can postdict AoA effects by putting AoA into the priors. At the same time, the model also explains masked priming effects via changes in the priors (Norris & Kinoshita, 2008). For example, the 50 ms presentation of the prime *table* is thought to temporarily increase the prior of the word target *TABLE,* leading to the prediction that *table* primes *TABLE*. A flexible approach to setting the priors ensures that the model can provide a good fit to a range of empirical data. Nevertheless, this ability to describe the data accurately comes at the cost of falsifiability.

A similar flexibility in the specification of the Bayesian Reader is evident in the choice of the likelihood function, which plays a critical role in the performance of the model. This function depends on assumptions about both the nature of input coding and the algorithm for matching this input representation against familiar words. Both of these components have undergone multiple iterations in the various papers describing the ongoing development of the Bayesian Reader. For example, the original model assumed the same coding scheme used in the original interactive activation model (McClelland & Rumelhart, 1981), but this prevented the model from accounting for a number of key priming results, including the masked priming obtained between primes and targets that differ in a letter-transposition (e.g., prime=*talbe,* target=*TABLE*; cf., Davis, 2006). In order to address this problem a subsequent version of the model, Norris and Kinoshita (2007) assumed the same coding scheme and matching algorithm used in the overlap model (Gomez, Perea, & Ratcliff, 2008). In a more recent paper, Norris et al. (2010) advocate a coding scheme similar to the spatial coding model (Davis, 2010) in order to recognize familiar words in novel contexts (e.g., CAT in the TREECAT), as well as various behavioral results that reveal the perceptual similarity of words that share a subset-superset relation (e.g., Bowers, Davis, & Hanley, 2005; Davis, Perea, & Acha, 2009; Grainger, Granier, Farioli, Van Assche, & van Heuven, 2006; cf., Bowers & Davis, 2009). This coding scheme has yet to be implemented in the Bayesian model.

We do not intend to criticize modelers for adopting a flexible approach that enables them to reject unsatisfactory solutions in favor of coding schemes and/or algorithms that work better. Nevertheless, this flexibility, which is characteristic of the Bayesian Reader, highlights the way in which critical aspects of Bayesian models are driven by the constraints provided by empirical data, rather than by rational analysis or fundamental Bayesian principles. Accordingly, statements like the following are not justified:

> *One of the most important features of the Bayesian Reader is that its behavior follows entirely from the requirement to make optimal decisions based on the available*

> *information ... The model has not been influenced or changed in any way by consideration of the data.* (Norris, 2006, p. 351).

For more discussion of these issues see Bowers (2010a, b) and Norris and Kinoshita (2010). For evidence regarding the "remarkable inefficiency" of visual word identification, see Pelli, Farell, and Moore (2003).

High-level cognition

High-level cognition would seem to be least amenable to a Bayesian interpretation given the vast catalogue of results suggesting that humans are poor (irrational) in making judgments, reasoning, and decision making. Kahneman and Tversky (1972, p. 450) concluded that, "In his evidence of evaluation, man is apparently not a conservative Bayesian; he is not a Bayesian at all." However, there has been a resurgence of research suggesting that we are near optimal in reasoning and decision making. But as above, these theories rely on post-hoc assumptions that weaken any conclusions. We briefly review three examples.

Decision-making in soccer. Bar-Eli, Azar, and Lurie (2009) consider the apparently irrational behavior of elite soccer players during penalty kicks. Statistical analysis shows that goalkeepers are most successful in blocking penalty kicks when they wait to see in which direction the ball has been kicked before choosing which way to jump. Nevertheless, the vast majority of the time, goalkeepers' jumps anticipate the kick. In addition, Bar-Eli et al. find that kickers are most likely to succeed when they shoot to the upper left or right of the net, and nevertheless, kickers most often shoot to the bottom half of the net. Bar-Eli et al. note that these choices seem irrational, particularly given that goalkeepers and kickers are highly trained athletes with large financial incentives to succeed. Why, then, do they not respond optimally to the statistics of penalty kicks?

Bar-Eli et al. (2009) answer this question by arguing that goal keepers and kickers are in fact responding optimally, but that the utility function of goalkeepers is not to minimize goals, and the utility function of kickers is not to maximize goals. Instead, their behavior is designed to optimize a "social rationality" utility function. In the case of a goalkeeper, the utility function includes not only the outcome (goal vs. no goal), but also his or her reaction to the outcome. That is, a goalkeeper might feel worse following a goal when he/she did not jump than when he/she did jump. This is not implausible: a goalkeeper wants to appear as though he/she is trying hard to stop the ball; trying your best is an important value. Similarly, the utility function of the kicker also includes his/her reaction to the outcome (goal vs. non-goal). A kicker is likely to feel worse following a missed goal if he/she missed the net compared to when the goal was saved by the goalkeeper. In the former case, the missed goal is likely to be entirely attributed to the failure of the kicker, whereas in the latter case, the failure will be attributed, to some extent, to the skills of the goalkeeper. So in both cases, the behavior of goalkeepers and kickers does not maximize the likelihood of saving/scoring goals but may maximize the social rationality utility function.

This seems a reasonable analysis, but the flexibility of the utility function undermines the strength of the conclusions. That is, if the results had been otherwise, and goalkeepers and kickers behaved in a fashion that maximized the utility function of goal outcome, then the same conclusion would obtain, namely, that soccer players act in an optimal fashion, just with different utility functions. Another example in which different utility functions lead to radically different outcomes is discussed by Nelson (2009), who notes the consequent difficulties in model comparison when "optimal" models can be outperformed by nonoptimal strategies.

The Wason card sorting task. A classic example of poor human reasoning comes from the Wason card sorting task. In the standard example, participants see cards with an A, K, 2, and

7 face-up, and are asked which cards they should turn over in order to test the hypothesis that "if there is an A on one side of a card, then there is a 2 on the other". Participants often turn over the card with the 2 (confirming the consequent), which is irrelevant, and rarely select the 7, which is relevant. This finding has been taken as strong evidence that people do not reason according to the rules of logic, and more generally, highlights our poor reasoning ability.

However, Oaksford, and Chater (1994) argue that performance is close to optimal when the normative standard is Bayesian rationality rather than logic, and they explain human performance with their *optimal data selection (ODS)* model. Very briefly, this model treats the Wason task as a conditional reasoning task. The question that participants are answering is whether there is a conditional relation between the antecedent (p) and the consequent (q), such that p(q|p) is higher than p(p) or p(q), or alternatively, whether (p) and (q) are independent, such that p(q|p) similar to p(p) or p(q). Based on this theory, participants turn over the cards that provide the most information in distinguishing these two hypotheses. Critically, the model can account for the pattern of results in this task if a rarity assumption is adopted – namely, that p (the card with the A in this example) and q (the card with the 2) are rare in the world. Under these conditions, the ODS model predicts that the two cards that lead to the greatest expected information gain are A and 2 – precisely the cards that subjects most often select.

There are reasons, however, to question this analysis. First, as noted by Evans and Over (2004) and Sloman and Fernbach (2008), the A and the 2 are not rare in the context of the cards set before the participants, and accordingly, it is not clear why participants should adopt the rarity assumption in this context. Second, the rarity assumption needs to be rejected in related cases. For example, Schroyens and Schaeken (2003) reported a meta-analysis of 65 experiments in which participants were asked to judge the validity of conditional inferences. For instance, a participant might be given the problem:

(1) If A then 2,

(2) A,

(3) 2?

That is, participants were required to determine whether or not the answer in (3) follows from (1) and (2). To fit these data with the ODS model it was necessary to assume that the probabilities of the antecedent and consequent were relatively high, contradicting the rarity assumption. That is, the rarity assumption must be invoked to explain performance on the Wason task, but must be rejected to explain conditional inferences in other tasks. Third, as noted by Wagenmakers (2009), each card in the Wason task reduces the expected uncertainty to some extent in the ODS model. Why, then, do participants not select all four cards rather than one or two?

Oaksford and Chater (2007) argue that their model succeeds when various pragmatic considerations are taken into account. With regards to the rarity assumption, they argue that judging the validity of conditional inferences (as in the studies reviewed by Schroyens and Schaeken, 2003) and selecting the relevant cards in order to test the validity of conditional inferences (as in the Wason card task) provide different contexts for interpreting the conditionals: That is, according to Oaksford and Chater, the conditionals are introduced as *assertions* in the former case, *conjecture* in the later case. These contexts are thought to encourage participants to reject and accept the rarity assumption, respectively. Oaksford and Chater also note that the pragmatics of the situation led the participants to pick some but not all the cards. One can agree or disagree with the force of the responses, but what is clear is that Bayesian principles *per se* do not account for human performance in these conditional reasoning tasks. Rather, ancillary assumptions need to be added to the model in response to the behavioral data.

Indeed, Oaksford and Chater (2007) agree that it is necessary to complement their computational level theory with a theory at the algorithmic level that incorporates various processing assumptions. Furthermore, their response to Schroyens and Schaeken's (2003) critique indicates that the algorithmic level need not simply be an implementation of the computational theory, but may contribute explanatory power in its own right. Oaksford and Chater (2003, p.151) write:

> *Oaksford et al. (2000) presented an account of conditional reasoning in which the computational level theory is provided by Bayesian probability theory. Consequently, Schroyens and Schaeken's single predictive failure only hits its target on the auxiliary assumption that the algorithmic level has no work to do. Oaksford et al. (2000) carefully avoided this assumption… Consequently, the one explanatory failure Schroyens and Schaeken found… does not mean that CP [the conditional probability model] is false…*

This response raises separate questions, however. If explanatory failures in a Bayesian model can be minimized by attributing the performance in question to processes occurring at the algorithmic level and outside the scope of the computational theory, is it ever possible to falsify Bayesian models, situated at the computational level? And if the behavior of the model may differ when the specific processing assumptions of the algorithmic level are implemented, how much weight should be attached to the successes of the computational theory? Might not these "correct" predictions be modified when the algorithm is implemented?

In sum, we have outlined several examples where priors, likelihoods, utility functions, and other considerations were introduced in response to the data (we review further examples in the next section). As noted by Sloman and Fernbach (2008), Bayesian models developed in this manner should be considered descriptive, not rational. Indeed, given the flexibility of these models to describe a wide range of outcomes, their successes provide little evidence that the mind computes in a Bayesian-like way, and furthermore, it undermines the use of these models as benchmarks for optimal behavior. If theorists want to claim that behavior approaches optimal performance, then it needs to be shown that Bayesian models capture human performance when the models were developed in response to the problem that needs to be solved, the environment, and perhaps some additional independent constraints derived from processing limitations (e.g., short-term memory) and biology—that is, when the priors, likelihood functions, etc., are developed without regard to the data themselves.

Falsification in Bayesian and non-Bayesian theories.

One possible reaction to the above critique is that it is unfair, because much the same criticisms could be levelled at all models. Non-Bayesian models often include many free parameters, and accordingly, they too can fit a wide range of potential outcomes, including those that were not observed. Thus, it could be argued that our criticisms concerning model falsifiability should be directed at modelling in cognitive science full stop, and not at Bayesian theories per se.

We have two general responses to this point. First, we agree that many non-Bayesian models are hard to falsify because they can account for a wide range of possible outcomes (including outcomes that do not in fact occur). Indeed, the flexibility of many standard processing models has been the focus of much discussion (e.g., Pitt, Myung, & Zhang, 2002; Roberts & Pasher, 2000). Nevertheless, 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. For instance, when comparing a Bayesian model of spoken word identification to previous non-Bayesian models, Norris and McQueen (2008) argue that the Bayesian model includes many fewer free parameters, writing:

> *So, where have all the parameters gone? Remember that our claim here is that people approximate optimal Bayesian recognizers, and that this determines the functions that must be computed. .. What allows us to dispense with so many free parameters are the strong principles underlying the theory.* (pp.387-388)

Or Geisler (2011) writes:

> *Thus, a powerful research strategy is to use the ideal observer to guide the generation of hypotheses and models of real performance… Models generated this way are principled and often have very few free parameters. (*p. 772)

But this is a mischaracterization of the flexibility of Bayesian models, as highlighted above. The choice of prior (as in motion perception), the likelihood function (as in the various versions of the Bayesian reader), the utility function (as in soccer players), and the alternative hypotheses that are considered (as in the Wason card sorting task) are all free to vary, and this is often exploited in order that the Bayesian model matches human performance. This weakens the value of a Bayesian approach both as a methodological tool and as a theory of mind.

Our second point is that Bayesian models are less easily falsified than non-Bayesian models because they generally make fewer predictions. That is, models developed at a computational level do not make predictions about phenomena that can be directly attributed to the algorithmic or implementational levels of analysis. For instance, Oaksford and Chater (2007) write:

> *There will clearly be aspects of cognition for which a rational explanation is not appropriate – the detailed cognitive effects of brain lesion can hardly be predicted from*

> *rational principles alone, after all – knowledge of the algorithms that the brain uses and their neural implementation will, of course, be crucial here. (p. 269).*

Or as noted by Griffiths, Kemp, and Tenenbaum (2008):

> *Some phenomena will surely be more satisfying to address at an algorithmic or neuro-computational level. For example, that a certain behavior takes people an average of 450 milliseconds to produce, measured from the onset of a visual stimulus, or that this reaction time increases when the stimulus is moved to a different part of the visual field or decreases when the same information content is presented auditorily, are not facts that a rational computational theory is likely to predict (p. 3).*

Although neuropsychological and response time data are often not considered relevant when developing Bayesian models at a computational level, these data are considered critical in the development and evaluation of standard information processing models. It stands to reason that it is harder to falsify the Bayesian accounts. Again, this weakens the strong conclusions often advanced by Methodological and Theoretical Bayesians.

Do Bayesian theories provide unique insight into the WHY question?

The flexibility of Bayesian theories also compromises what is often considered to be one of their key advantages, which is that they are claimed to provide unique insight into the *why* question. In fact, there are two different ways in which theorists have asserted the advantage of Bayesian approaches in answering the *why* question. The first, more common, argument is that computational-level Bayesian theories are able to explain all of the data of interest without invoking arbitrary mechanistic explanations. For example, Oaksford and Chater (2009) write:

> *...a good deal of empirical data about human reasoning (and indeed, human cognition more generally) can be understood as arising from the structure of the problem itself –*

> *that is, the nature of the problem drives any reasonable algorithmic solution to have particular properties, which may be evident in the data. This idea is a core motivation for the rational analysis approach... (p. 84).*

Regarding this approach, if behavior approximates the predictions that can be derived from a purely rational analysis, it provides an answer to the *why* question: Behavior is as it is because it is close to optimal. For present purposes, we will call this the Type-1 answer to the *why* question.

A second, rather different argument starts from apparent deviations from optimality. The assumption is that performance is in fact optimal, despite appearances, and the goal is to identify assumptions (e.g., priors, likelihoods) that make the observed behavior (near) optimal. That is, theorists should compare different Bayesian models to the data (e.g., models that make different assumptions about the mind and world), and the success of a particular Bayesian model is thought to provide insight into what assumptions the mind works with. The critical point for present purposes is that these assumptions play a key role in answering the *why* question: People act in a certain way because they are rational and have specific assumptions about the world. This view is nicely captured by Griffiths and Tenenbaum (2006, p. 772), who write:

> *Bayesian models require making the assumptions of a learner explicit. By exploring the implications of different assumptions, it becomes possible to explain many of the interesting and apparently inexplicable aspects of human reasoning.*

For present purposes, we refer to such assumptions as Type-2 answers to the *why* question.

Although both types of answers to the *why* question have been characterized as Bayesian, they are quite different, most notably with respect to whether the data themselves are supposed to contribute to the development of explanations. In our view, neither approach provides a reason to prefer Bayesian over non-Bayesian theorizing.

The key problem with the Type-1 answers is that Bayesian models are rarely if ever constructed on the basis of rational analysis alone. Rather, as demonstrated above, modellers are free to vary many aspects of their model, and they often need to make arbitrary assumptions about priors, likelihood functions, etc. in order to capture the data. A rational analysis informed by human performance is every bit as arbitrary as a mechanistic model built around the data, and the predictions of such models should be considered descriptive rather than normative. As descriptive theories, they lose their capacity to provide Type 1 answers to the *why* question.

Type-2 answers are more subtle. A good illustration of this approach is offered by McKenzie (2003), who describes a number of instances in which performance in the laboratory appears to deviate from optimality but where a rational model with the appropriate assumptions can explain behavior. One of his examples is the so-called “framing effect” in which logically equivalent descriptions of a problem lead to very different decisions (Tversky & Kahneman, 1981). Consider the standard example of evaluating a medical treatment. Participants are told to imagine that they have a terrible disease and have to decide whether to accept a specific treatment option. Some participants are told that the treatment has a 20% mortality rate within 5 years while other participants are told that the treatment has an 80% survival rate after 5 years. A reliable finding is that participants are more likely to accept the treatment described in “survival” terms (e.g., Marteau, 1989; Wilson et al., 1987). This would seem to be irrational given that the two problems are logically equivalent. However, McKenzie (2003) argues that framing effects are rational when the appropriate assumptions are added to the analysis. That is, in the real world, the way a problem is framed is often informative—it reflects knowledge of the person who posed the problem—and accordingly, it should be given some weight. If, for example, a doctor frames a treatment in a positive light (it has an 80% chance of success), he or she is implicitly suggesting that it is a good option, and should be taken. According to McKenzie, people often pick up on this information in the real world and mistakenly use the same strategy in a laboratory context in which framing cues should (logically) be given no weight. So, according to McKenzie, the Type-2 answer to the *why* question is that it is rational in the real world for choices to be influenced by the framing of a problem, and participants take this assumption with them into the laboratory.

We see two limitations of this latter approach. First, it is subject to the same concerns of too much flexibility – it will always be possible to find some, perhaps many different, assumptions that will allow a Bayesian model to accommodate human performance. Indeed, as noted by Jones and Love (2011), Bayesian theorists rarely compare alternative Bayesian models (that adopt different assumptions), and accordingly, some quite different set of assumptions may provide an equally good or even better Type-2 answer. Second, and more important, there is nothing unique or novel about Bayesian Type-2 answers to the *why* question. The most obvious parallel is with the explanations offered by evolutionary psychologists. Theories in this discipline attempt to explain irrational or surprising behavior in an adaptive manner by looking for assumptions that the brain carries from our evolutionary past. For example, there are a number of studies showing that mutual exposure during childhood weakens sexual attraction among adults. Why is this? It might at first be expected that familiarity and a common history would increase the likelihood of mutual attraction. The answer from evolutionary psychology is that it is important to avoid inbreeding, and accordingly, we have evolved a system for detecting genetic relatedness that prevents us from being sexually attracted to persons we grew up with (e.g., Lieberman, Tooby, & Cosmides, 2007). The Type-2 answers advanced in evolutionary psychology are often criticized (sometime fairly, sometimes not in our view) as “just-so” stories, and we do not see any way in which Bayesian theories are more immune to these criticisms.

Part 3: Bayesian theories are rarely compared to alternative (non-Bayesian) hypotheses

Theoretical Bayesians make a stronger claim than Methodological Bayesians, and accordingly, their models require stronger evidence. In particular, in order to provide some evidence in support of the claim that the mind relies on Bayesian-like algorithms, it is necessary to show that these algorithms do a better job than non-Bayesian models in accounting for human performance. However, this is rarely done. Instead, Theoretical Bayesians generally take the successful predictions of a Bayesian model as support for their approach, ignoring the fact that alternative theories might provide an equally good fit to the data. Below we review a number of examples from different domains in which evidence taken to support Bayesian theories could also be interpreted as support for non-Bayesian heuristic theories.

Optimal predictions in everyday cognition

As noted already, there is a vast literature highlighting the many conditions in which people are poor at reasoning, judgments, and decision making (compared to some normative standard). Given this context, Griffiths and Tenenbaum’s (2006) claim that people make optimal predictions in everyday cognition is surprising. Their claim is based on the results of an experiment in which students were asked to make predictions about the duration or extent of 8 different phenomena given partial information as described in a short passage. For example, one question was as follows, with the numbers (queries) in brackets varying for different groups of students:

> *Insurance agencies employ actuaries to make predictions about people's life spans—the age at which they will die— based upon demographic information. If you were assessing an insurance case for an [18, 39, 61, 83, 96] year-old man, what would you predict for his life span?*

In each of these scenarios Griffiths and Tenenbaum were able to develop an optimal Bayesian estimate given the query (e.g., the current age of the person in the above question), an empirically derived prior probability (e.g., the distribution of actual life expectancy), a likelihood function (assumed to be a uniform function between 0 and the participant's prediction), and a decision criterion (the median of the posterior probability function). The critical finding was that the students' predictions often matched the optimal prediction. This result was taken to suggest that their judgments were informed by prior distributions that are accurately calibrated to the statistics of relevant events in the world, and further, that these priors are utilized in Bayesian computations in order to make nearly optimal decisions. Griffiths and Tenenbaum concluded: "These results are inconsistent with claims that cognitive judgments are based on non-Bayesian heuristics that are insensitive to priors" (p. 770).

In order to support the conclusion that judgments in this study were the product of Bayesian processes informed by detailed and accurate priors, it is necessary to show that alternative non-Bayesian models are not just as good. Accordingly, it is significant that Mozer, Pashler, and Homaei (2008) developed a simple heuristic model that accounted for the data just as well on the basis of highly impoverished knowledge of the relevant statistics of the world: each decision was based on a sum total of two memories rather than detailed and accurate priors. This model was not intended to provide a psychologically plausible theory of decision making, but was rather developed to show that the strong conclusions of Griffiths and Tenenbaum (2006) were unjustified. It succeeded in this respect. Subsequently, Lewandowky, Griffiths, and Kalish (2009) showed that the Bayesian model fared better than Mozer et al.'s heuristic model in accommodating the results of a new set of experiments, but they still did not attempt to contrast their hypothesis with plausible alternative theories.

In fact, it is not hard to imagine that a non-Bayesian theory could account for performance in this task. For example, in response to the lifespan question, it seems plausible that the students did not rely on memories of dead individuals at all, but instead, know (as a fact) that life expectancy is about 75, and discount as largely irrelevant the information that the person in the query is currently 18, 39, or 61 year old. But when the person in question is 83 or 96, the participants necessarily know that the person will live over 75, and presumably some extra time beyond their current age (that is, it is unlikely that the person is interviewed on the day of their death). In short, a serious examination of alternative hypotheses is needed before the strong conclusions of Griffiths and Tenenbaum (2006) and Lewandowsky et al. (2009) are accepted. This is particularly important in the current context given that a wide range of findings in the field of behavioral economics suggest that human judgment and decision making is often quite poor (cf., Kahneman, Tversky, 1996; Kahneman, 2003). That is, there is a low prior probability that humans make optimal predictions in everyday life, and so strong evidence is required to support a conclusion to the contrary.

Memory

The hypothesis that human memory is close to optimal might also appear to be surprising given its many limitations, including our tendency to forget, and the frequency with which we misremember information. Schacter (2001) notes seven ways that memory seems to let us down.

Nevertheless, it is commonly claimed that human memory can be characterized as exemplifying optimal Bayesian inference.

Xu and Griffiths (2010) recently developed a Bayesian model of serial reproduction memory in which one person's recall (reconstruction) of a stimulus is presented to the next person as the to-be-remembered stimulus. In a behavioral study, the participants were trained to distinguish images of two categories of fish: simple drawings of farm fish and ocean fish. The critical manipulation was that the distribution of the sizes of the farm fish varied across two groups, such that the participants in these two groups were expected to learn different priors for the width of farm fish. In the reproduction phase, a target farm fish stimulus was flashed for 500 ms and disappeared, followed by probe fish of a random size. Participants had to adjust the width of the probe fish to match the target. The first participant in each group saw a target fish of variable width, and subsequent participants were presented with the reproductions. The critical finding was that performance of the two groups diverged over iterations, such that recall (i.e., the ultimate adjusted width) in the later iterations was strongly influenced by the priors. This work extended previous findings of Huttenlocher, Hedges, and Vevea (2000), and provided a formal replication of the classic work of Bartlett (1932).

Xu and Griffiths (2010) provided a rational analysis account of this pattern of results. In this view, a memory system should attempt to store accurate information about the world given noisy sensory data. That is, during perception, people seek to recover the true state of the world that generated the noisy stimulus and then store this estimate in memory. Memory retrieval is characterized as retrieving this estimate (rather than the noisy sensory data), and in the context of this experiment, the retrieved memory serves as the stimulus for the next participant. Xu and Griffiths show that such an iterative process in a Bayesian model results in memory retrieval that converges on the model's prior, consistent with the data. This is taken to support their model. The problem with this conclusion, however, is that non-Bayesian theories could account for this pattern of results as well. As long as memory of an event is biased towards pre-existing knowledge, it is likely that memory in a serial reproduction task will converge to this knowledge; errors will tend to fall in the direction of the bias. That is, these data do not distinguish Bayesian from non-Bayesian theories, unless Bayesian theories are trivialized to the point of meaning that memory is biased by pre-existing knowledge.

Similar concerns apply to a Bayesian model of memory for object size proposed by Hemmer and Steyvers (2009). The authors showed that memory for the size of an object was influenced by priors (e.g., prior knowledge that strawberries tend to be small influenced memory for the size of strawberries), and the authors took this to support a Bayesian theory of memory. But again, a non-Bayesian theory that assumes that memory is biased by pre-existing knowledge would likely be able to accommodate this general trend as well, as errors would tend toward the bias. Furthermore, in order to fit the data successfully, it was necessary to add more noise to the memory process under some conditions (i.e., when retrieving unfamiliar objects within a familiar category). Although this choice of likelihood function was successful for Hemmer and Steyvers' data, it is not clear how adding noise to the memory process is compatible with other findings, such as the finding that recognition memory is better for low compared to high frequency words (Balota & Neely, 1980). If more noise is added to the process of retrieving less familiar things, then the obvious prediction would be that recognition memory is worse for low frequency words, which is contrary to fact. In order to account for the memory advantage of low compared to high-frequency words, Shiffrin and Steyvers (1997) developed a Bayesian model of recognition memory in which high frequency words were more similar to one another compared to low frequency words. They did not provide any motivation for this assumption in their model other than to account for the effect. If this assumption were applied to the Hemmer and Steyvers (2009) context, then it is no longer clear that the modified model could account for their results. In short, the successes of these Bayesian models of memory are either quite general (in the sense that non-Bayesian models would likely succeed as well) or have little to do with their Bayesian principles.

Perhaps the most salient failure of memory is our tendency to forget. Anderson and Schooler (1991) advanced a Bayesian theory that explains why it is optimal to forget. In this view, given the vast amount of information that we have encountered and the limited resources of the mind to retrieve information, the adaptive thing is to forget things we are unlikely to need in the future. The optimal form of forgetting is one in which memories that are most likely to be needed are retained and memories least likely to be needed are forgotten. Critically, Anderson and Schooler reported evidence for just this pattern of forgetting. They analyzed the probability of a word being included in the front page headlines in the *New York Times* as a function of the number of days since the word was last included. They observed that word usage followed a power function over time, and this was taken to characterize the likelihood that someone will need to remember something. Consistent with these results, various studies have reported that episodic memory deteriorates as a power function of the interval (Rubin & Wenzel, 1996; Wixted & Ebbensen, 1991; but see Wixted & Carpenter, 2007). Anderson and Schooler (1991) took this as evidence that forgetting is optimally suited for the environment we live in.

There are some problems with this conclusion, however. First, as noted by Becker (1991), headlines in newspapers are the products of the human mind. Accordingly, it is not surprising that an environment produced by the human mind parallels human memory performance; indeed, in this case, forgetting may not conform to the nature of the environment, but instead, the environment conforms to forgetting. Second, non-Bayesian models can account for power function forgetting curves (e.g., Sikstrom, 1999). Accordingly, this finding alone provides no evidence for Bayesian theories. Third, and perhaps most importantly, forgetting can dramatically deviate from a power function over time, as in retrograde amnesia (where older episodic memories are better preserved than more recent ones; cf. McClelland, McNaughton, & O'Reilly, 1995) or *age-of-acquisition* (AoA) effects, where early acquired aspects of language show an advantage over later acquired aspects (e.g., Juhasz & Rayner, 2006; Stadthagen-Gonzalez et al., 2004; Bowers, Mattys, & Gage, 2009; Scherag, Demuth, Roesler, Neville, & Roeder, 2004; Johnson & Newport, 1989). These phenomena have been modeled in non-Bayesian models that derive their predictions from processing constraints at the algorithmic level (e.g., McClelland et al., 1995; Ellis & Lambon Ralph, 2000; McClelland, 2006). It is unclear how a Bayesian theory could provide a principled account of these deviations from a power function.

Motor control

A large literature has highlighted how motor control is close to optimal in a variety of contexts (cf., Wolpert, 2007). For example, Trommershauser, Maloney, and Landy (2003) asked participants to touch a briefly displayed green target region on a computer monitor while avoiding one or more partially overlapping red regions. The experiment manipulated the relative benefits and costs of touching green and red regions. The challenge for the participants, then, was to strike an appropriate balance between the goal of hitting the target and the goal of avoiding the penalty region. Trommershauser et al. (2003) then compared the performance of the participants with a model that included the same variability in responding but was designed to maximize the mean expected gain. In most cases it was not possible to discriminate the performance of the participants from the model, leading Trommershauser et al. to conclude that human motor planning is close to optimal.

However, Wu, Trommershauser, Maloney, and Landy (2006) reached a different conclusion when comparing a Bayesian to a non-Bayesian account of motor control in more complex environments. In the Trommershauser et al. (2003) study the configuration of the goal and penalty areas ensured that the point of mean expected gain always fell on an evident axis of geometric symmetry, which reduced the difficulty of the task. Wu et al. note that if the participants used the axis of symmetry to constrain responding, then the behavior might be better characterized as a “motor heuristic” rather than an optimal Bayesian strategy. In order to contrast Bayesian and heuristic approaches, Wu et al. developed more complex arrangements of goal and penalty areas, such that the point that maximized mean expected gain fell outside an evident axis of symmetry. As predicted by a heuristic account, performance fell off dramatically when the point of maximal expected gain fell far from the evident axis of symmetry.

In contrast with this work, most studies have not evaluated Bayesian motor control theories in the context of alternative plausible theories. For example, in a classic study taken to highlight the Bayesian nature of motor control, Kording and Wolpert (2004) asked participants to point to a target when given distorted visual feedback via a virtual reality setup. Specifically, for each movement there was a lateral shift randomly drawn from a prior distribution with a mean shift of 1 cm to the right, and the feedback was either provided clearly or blurred to increase the uncertainty of the feedback. Critically, participants took into account both the distribution of lateral shifts and the clarity of the feedback in order to respond in close to an optimal manner, consistent with a Bayesian model. Kording and Wolpert (2004) did compare human performance to two alternative models that did not incorporate any information about the prior distribution or visual clarity of the feedback, and found that the Bayesian model provided a better fit to the data.

But it is hardly surprising that participants learn something about the distributions of displacements during training and use this knowledge to inform their responses, and further, that they are sensitive to the quality of the feedback on a given trial. The question is whether plausible alternative heuristic models can account for behavior, but this was not tested. Nevertheless, based on the relative success of their Bayesian model, Kording and Wolpert (2004) suggested that human performance might be optimal in a wide range of conditions:

> *Although we have shown only the use of a prior in learning hand trajectories during a vasomotor displacement, we expect that such Bayesian process might be fundamental to all aspects of sensorimotor control and learning. (p. 246)*

These strong conclusions are difficult to reconcile with a growing set of behavioral studies that have reported striking suboptimalities in motor control. For example, Zhan, Wu, and Maloney (2010) found that participants performed suboptimally on a task that involved touching two targets in a specified order, even after extensive training. Given the common failure to contrast Bayesian and heuristic accounts of motor control (for a nice illustration of the contrast between Bayesian and non-Bayesian theories of motor control, see Gigerenzer and Selten, 2001), and recent demonstrations of suboptimal performance (e.g., Burr, Banks, & Morrone, 2009; Mamassian, 2008; Wu et al., 2009; Zhan et al., 2010), we conclude that there is little evidence that motor behavior is mediated by Bayesian as opposed to non-Bayesian process.

Multi-Sensory Perception

Many studies have highlighted how our perceptual systems are near optimal in combining inputs from multiple modalities in order to estimate properties of the world. In order to combine evidence in an optimal manner, the relative noise (uncertainty) associated with the inputs needs to be considered so that greater weight can be given to more reliable inputs. Various behavioral studies have demonstrated that human performance in such conditions is extremely similar to a model that optimally combines information from different modalities (e.g., Alais & Burr, 2004; Ernst & Banks, 2002). For example, Ernst and Banks (2002) asked participants to look and/or feel a raised ridge and judge its height. The authors measured the accuracy of each modality alone by asking participants to carry out discrimination experiments (which ridge was taller) in either the visual or haptic modality alone. In the visual modality the reliability of the input was varied across four conditions, with various amounts of visual noise added. In the joint condition, participants made further discriminations when the visual and haptic information was combined. In some trials the visual and haptic information matched, and in other trials, the visual and haptically specified heights of the ridges differed. The estimates of the reliability of the visual information (that varied with noise) and haptic information alone provided estimates of optimal performance in the multi-modal condition, and human performance was "remarkably similar" (p. 431) to optimal in the combined condition. After reviewing the literature on integrating perceptual information both within and between modalities, Kording and Wolpert (2006) concluded that:

> *…people exhibit approximately Bayes-optimal behaviour. This makes it likely that the algorithm implemented by the [central nervous system] may actually support mechanisms for these kinds of Bayesian computations (p.322).*

The problem with this conclusion, however, is that non-Bayesian algorithms can also combine cues in ways that approximate optimal performance (e.g., Gigerenzer & Brighton, 2009; Juslin, Nilsoon, & Winman, 2009), and despite the dozens of studies that have highlighted the near optimality of cue integration in multi-sensory perception, there has been little or no consideration of alternative non-Bayesian solutions. Accordingly, the conclusion of Kording and Wolpert (amongst others) is unwarranted.

In sum, given the computational complexity of implementing Bayesian computation in the brain, we would argue that the onus is on Theoretical Bayesians to show that models that do implement such computations can explain human performance better than non-Bayesian models. This, as far as we are aware, has never been demonstrated.

Non-Bayesian approaches sometimes provide a better account of human performance than Bayesian theories

As detailed above, Bayesian models are rarely compared to non-Bayesian models, and when they are, there is little evidence that Bayesian models perform better. Indeed, non-Bayesian models sometimes provide a better account of human performance, as we briefly review in the following example.

Consider the phenomenon of probability matching. When asked to make predictions about uncertain events, the probability with which observers predict a given alternative typically matches the probability of that event occurring. For instance, in a card guessing game in which the letter A is written on 70% of the cards in a deck and the letter B is written on 30%, people tend to predict A on 70% of trials (e.g., Stanovich, 1999; West & Stanovitch, 2003; for related examples, see Estes, 1964). This behavior is irrational, in the sense that it does not maximize utility. The optimal (Bayesian) strategy would be to predict the letter A every time, as this leads to an expected outcome of 70% correct, whereas perfect probability matching would predict an accuracy rate of only .70 * .70 + .30 * .30 = 58%.

Why would observers adopt a selection strategy that is suboptimal? There have been attempts to explain this pattern of results within a Bayesian framework. For instance, Wozny, Beierholm, and Shams (2010) take what we referred to above as a Type-2 approach to answering the *why* question. That is, the claim is that observers are near optimal but come to the task with the implicit (and incorrect) assumption that the sequence of experimental trials contains predictable patterns (see Gaissmaier & Schooler, 2008 for a related account). Under these conditions, Wozny et al. (2010) argue that it is optimal for the brain to sample from a distribution over hypotheses, where each hypothesis corresponds to a different causal model. Although such a sampling procedure would indeed give rise to the observed probability matching behavior, there are problems with the argument that probability matching reflects pattern seeking behavior. For instance, Koehler and James (2009) obtained no evidence that participants who showed probability matching were in fact looking for sequences: When the task was changed so that patterns could not in principle be exploited to support predictions, participants continued to probability match. Furthermore, this rational account of probability matching is hard to reconcile with the finding that a small subset of more intelligent participants rejected probability matching in favor of the optimal strategy (West & Stanovich, 2003). If it is rational to probability match, why are the most intelligent participants acting differently? Of course, it might be argued that the group of probability matchers and the group of optimizers are both Bayesian, but adopt different assumptions (i.e., the more clever participants do not look for hidden patterns). But this again highlights how Bayesian theories are always flexible and often built post-hoc around the data.

In our view, a more promising explanation of probability matching behavior of the majority of participants is offered by an adaptive network approach. For example, consider a simple neural network with one stimulus node and two response nodes corresponding to the two possible outcomes of the letter A or B. Assume that the response nodes laterally inhibit each other so that the network predicts either A or B. On any given trial, the winning node is the one that receives the largest noisy input, where the non-stochastic component of the input to a response node depends on the adaptable weight that connects it to the stimulus node. These weights z_1 and z_2 are updated by a simple learning rule that modifies the weights based on the accuracy of the network's predictions:

$$\Delta z_i = \alpha\ (o_i - y_i)\ x \qquad (5)$$

In this equation, the term $(o_i - y_i)$ is an error-correction reinforcement signal that reflects the difference between the observed outcome o_i (i.e., whether the card was an A or a B) and the predicted outcome y_i,; α is a learning rate parameter. Thus, learning occurs when the relevant stimulus is present (x=1 denotes the presence of a card) and the network's prediction is wrong. When learning does occur, the weight z_i decreases if the network incorrectly predicted outcome y_i or increases if this outcome occurred when it was not predicted. This simple learning rule, which is known as the perceptron learning rule (Rosenblatt, 1962) or the delta rule (Widrow & Hoff, 1960), is capable of learning a broad range of pattern discriminations. It can also be directly related to the classic Rescorla-Wagner (1972) model of associative learning,and can explain many of the well known aspects of associative learning.

For present purposes, the interesting aspect of this simple network is that its predictions perfectly mirror the observed probabilities, i.e., the mean response rate for y_i is asymptotically equal to the mean observed rate of o_i. This is true even when the number of possible outcomes is increased from two to an arbitrarily large value. The reason for this is that the network approximates a simple heuristic strategy called *win-stay, lose-shift* (Robbins, 1952), in which the individual tends to stay with an alternative that provides a reward, but switches randomly to another alternative when they do not receive a reward. Indeed, in the absence of input noise, the network follows this strategy exactly. The win-stay, lose-shift strategy leads to probability matching because the pattern of responses follows the pattern of observed outcomes, shifted by one trial. Note, the win-stay, lose-shift strategy is one that people frequently adopt (e.g., Steyvers, Lee, & Wagenmakers, 2009) despite the fact that such a strategy is profoundly non-Bayesian (in that it does not take into account the distribution of reward rates for each option).

In sum, an adaptive network account can provide a simple explanation of the apparently irrational probability matching behavior. Note that this approach does not attempt to store all of the instances that have been encountered; learning is influenced by the sampling distribution, but there is no explicit encoding of this distribution. It is also the case that no new assumptions were introduced in response to the data in order to account for the data: probability matching and the win-stay, lose-shift response pattern directly follow from the delta-rule that was introduced to account for unrelated phenomena in a biologically plausible manner. By contrast, the attempt to reconcile the observed behavior with an optimal Bayesian account, according to which the brain samples from a distribution over causal hypotheses on each trial, strikes us as ad hoc and unparsimonious. Although Bayesian calculations are optimal for updating the probabilities of events, much simpler adaptive learning algorithms are also able to learn such probabilities, and, critically, can explain why observers' exploitation of probability information is suboptimal (for similar arguments in another context, see Goldstein & Gigerenzer, 2002).

Confirmation Bias in Bayesian and non-Bayesian theories

The key point that we have tried to make in Part 3 is that advocates of Bayesian theories often show strong confirmation bias: Theorists take the successful predictions of a Bayesian model as support for their approach and ignore the fact that alternative non-Bayesian theories might account for the data just as well, and sometimes better. We take this to weaken the Theoretical Bayesian position according to which the mind computes in a Bayesian-like way, just as we took the flexibility of Bayesian models as compromising Methodological and Theoretical Bayesian positions in Part 2.

A possible objection to this argument is that confirmation bias is ubiquitous in science and that the criticisms outlined above apply to non-Bayesian theories just as well. However, we would suggest that this bias is more problematic in the Bayesian case. Bayesian theories always predict that performance on a task is *near* optimal, and accordingly, the predictions of these models are often similar to what one would expect from a satisficing (but non-Bayesian) solution. At first it seems impressive that Bayesian models can approximate performance in all variety of domains, from low-level perception to high-level word and object identification as well as motor control, memory, language, semantic memory, etc. But this impression should be tempered by the fact that alternative (adaptive) models would be likely to predict (or already have predicted) the findings just as well. What is needed is a method of distinguishing between Bayesian and non-Bayesian models that can both account for near optimal performance (cf., Maloney & Mamassian, 2009).

What would provide a powerful confirmation of a Theoretical Bayesian approach is an unexpected (counter-intuitive) prediction that is supported. In principle this could take the form of a model doing surprisingly well—better than what would be expected on the basis of heuristic models that satisfice as opposed to optimize. But we are not aware of any examples in which human performance is better than what could be accounted for with a heuristic model. Or alternatively, this could be achieved by accounting for performance that is surprisingly poor, worse than would be expected based on a heuristic approach. In fact, Bayesian models are often used to explain poor performance in various domains: As noted above, Bayesian models can account for forgetting in episodic memory, poor strategy in penalty kicking by elite soccer players, illusions of motion perception, etc. But in most (perhaps all) cases, the models succeed in doing badly by adding constraints in response to the data. Again, non-Bayesian models could account for these findings in the same way – by looking at the data and building in a set of constraints that allow the model to succeed in accounting for poor performance.

Where then are the opportunities to provide a strong test of a theory? The non-obvious and counter-intuitive predictions of a psychological model are typically derived from non-functional constraints found at the algorithmic level, a level of descriptions that most Bayesian models avoid. For example, Parallel Distributed Processing (PDP) models that learn dense distributed representations are good at generalizing but poor at learning information quickly because of so-called *catastrophic interference.* Accordingly, McClelland et al. (1995) argued that these two functions must be carried out in separate systems; the so-called complementary learning systems hypothesis. This is not a computational constraint, as alternative models can generalize and learn quickly within a single system (e.g., Grossberg, 1987). Accordingly, it is not clear how a rational analysis carried out at the computational level could shed light on this issue (cf., McClelland et al., 2010).

In sum, we do not intend to hold Bayesian and non-Bayesian models to different standards. We readily acknowledge there are plenty of examples in which non-Bayesian modelling can be characterized as a form of curve fitting. Confirmation bias is rife in all areas of psychology and science more generally. However, models developed at an algorithmic level often make strong (counterintuitive) predictions, whereas Bayesian models rarely do (though see Hahn & Warren, 2009, for a nice example in which a rational analysis does in fact lead to a counterintuitive prediction). As a consequence, Bayesian models tend to receive credit for relatively trivial predictions, which could be derived from any theory that assumes performance is adaptive, or else are modified with post-hoc assumptions when the data do not follow what would be predicted on the basis of the rational analysis alone.

Part 4: Neuroscientific evidence supporting Bayesian theories is weak

At the same time that Bayesian theories have become so prominent within psychology, Bayesian theories have become prominent in neuroscience. The general claim is that populations of neurons represent uncertainty in the form of probability distributions and perform (or approximate) Bayesian computations in order to make optimal decisions. This view is sometimes referred to as the Bayesian coding hypothesis (Knill & Pouget, 2004).

The Bayesian coding hypothesis is often taken as an additional source of evidence for Bayesian theories in psychology. For example, Chater, Tenenbaum, and Yuille (2006) write:

> *...turning to the implementational level, one may ask whether the brain itself should be viewed in probabilistic terms. Intriguingly, many of the sophisticated probabilistic models that have been developed with cognitive processes in mind map naturally onto highly distributed, autonomous, and parallel computational architectures, which seem to capture the qualitative features of neural architecture (p. 290).*

If neuroscience provided evidence for the Bayesian coding hypothesis, it would indeed provide support for the Theoretical Bayesian perspective. The problem, though, is that advocates of the Bayesian coding hypothesis often cite the psychological literature as providing the best evidence for their hypothesis. For instance, Knill and Pouget (2004, p.712) write that “the principle [sic] data on the Bayesian coding hypothesis are behavioral results showing the many different ways in which humans perform as Bayesian observers”.

If neuroscience is to provide any evidence for the Theoretical Bayesian perspective, the key question is what non-behavioral evidence exists that neurons compute in this way? The answer is none, unless Bayesian computations are characterized so loosely that they are consistent with almost any computational theory.

Does neural noise necessitate Bayesian inference?

A common starting point for a Bayesian characterization of neural computation is that neural firing is very noisy, with the same input generating quite different neural responses (e,g., Faisal, Selen, & Wolpert, 2008). This makes the task of a neural network statistical, in that the network needs to infer something fixed about the world or devise a specific motor plan based on noisy (ambiguous) input.

A key insight of Pouget and colleagues (e.g., Ma, Beck, Latham, & Pouget, 2006) is that if neurons show near-Poisson variability, then Bayesian inference can be implemented through a simple linear combination of populations of neural activity. A fundamental property of a Poisson response is that the variance and the mean are the same, and a number of experimenters have found the ratio of the variance to the mean spike count to range from about 1-1.5 (e.g., Shadlen & Newsome, 1998) – the so-called Fano factor. Indeed, this correspondence has been described as “remarkable” (Beck et al., 2009, p. 6). It is not only taken as an existence proof that Bayesian computations are possible by neurons but also as evidence in support of the Bayesian coding hypothesis (e.g., Beck et al., 2008; Jazayeri, 2008).

However, is also important to note that Fano factors are often much smaller than one (e.g., Amarasingham, Chen, Geman, Harrision & Sheinberg, 2006; Maimon & Assad, 2009), and sometimes approach zero (e.g., Gur, Beylin, & Snodderly, 1997; Gur & Snodderly, 2006; DeWeese, Wehr, & Zador, 2003; Kara, Reinagel, & Reid, 2000), reflecting a remarkable lack of noise in the nervous system. For example, DeWeese et al. (2003) assessed the trial-by-trial response variability of auditory neurons in the cortex of rats in response to tones. The reliability of individual neurons was almost perfect, leading DeWeese et al. to suggest the need for a "re-evaluation of models of cortical processing that assume noisiness to be an inevitable feature of cortical codes". (p., 7490) Similarly, on the basis of single-cell recording from neurons in the inferior colliculus (IC) of guinea pigs, Shackleton, Skottun, Arnott, and Palmer (2003) concluded that "any pooling or comparison of activity across a population is performed for reasons other than reducing noise to improve discrimination performance". (p. 723) For a recent review highlighting the reliability of single cell responses and an argument that information is often represented by the firing of single ("grandmother") neurons, see Bowers (2009).

Of course even if single neurons are more reliable than assumed by Pouget and colleagues, this does not rule out the claim that collections of neurons compute in a Bayesian-like fashion. Still, these considerations challenge one of the common motivations for the Bayesian coding hypothesis: the idea that noise that "permeates every level of the nervous system"*;* Faisal et al., 2008, p. 292) as well as challenging one piece of evidence often taken to support this view , that firing variability is close to Poisson in order to facilitate Bayesian computations.

What is the evidence for the Bayesian coding hypothesis?

Mathematical analyses have shown how Bayesian inferences can be implemented in populations of neurons given certain types of neural noise. The key question, though, is whether there is evidence that neurons do in fact compute in this way. We would suggest that the current evidence is either poor or inconsistent with the hypothesis. We review the most relevant evidence below.

Ma and Pouget (2008) considered how multisensory neurons should optimally combine inputs from different modalities. A prediction of the Bayesian coding hypothesis developed by Pouget and colleagues is that multisensory neurons should fire in an additive fashion, such that their firing rate in response to two inputs should equal the sum of their responses when each input is presented separately, under on the assumption of Poisson-like variability. However, because firing rates of neurons saturate as they approach their maximum rate, firing rates of multisensory neurons can be subadditive as well (which, under appropriate conditions, will not affect optimality). That is, according to Ma and Pouget (2008) a Bayesian theory predicts additive or subadditive firing rates. Consistent with this prediction, Ma and Pouget note that the majority of multisensory neurons in the cat superior colliculus exhibit additivity or subadditivity.

What is problematic for this theory, however, is that multisensory neurons often respond in a superadditive manner as well. Indeed, as Ma and Pouget (2008) note, superadditivity is often taken as evidence that a particular neuron is involved in multisensory integration. Although Ma and Pouget suggest various ways in which a Bayesian theory might potentially account for superadditivity, the key point is that multisensory neurons respond in an additive, subadditive, and superadditive fashion. Given that all possible effects have been obtained, these results should not be taken as evidence in support of a Bayesian account of multisensory perception.

Another argument advanced in support of the Bayesian coding hypothesis was summarized by Jazayeri (2008). He reviewed a wide range of findings in which two measures of neuronal behavior are found to be correlated, namely, “sensitivity” and “choice probability”. Sensitivity refers to the reliability with which a neuron fires to a given stimulus, whereas choice probability refers to its ability to predict trial-to-trial behavior in a choice task. For example, Britten, Newsome, Shadlen, Celebrini, and Movshon (1996) trained monkeys to discriminate between two directions of motion and found that the more sensitive the neuron, the better able it was to predict the monkey’s response. Similar correlations have been reported in various tasks, from visual feature discrimination tasks (Purushothaman & Bradeley, 2005), perception of time tasks (Masse & Cook, 2008), heading discrimination (Gu, DeAngelis, & Angelaki, 2007), amongst others. Together, these findings are taken as providing evidence that the brain combines and weights sensory signals according to their reliability, with more sensitive neurons given more weight, consistent with Bayesian principles.

But this general pattern of results should be expected by any theory, Bayesian or not. The correlation of sensitivity and choice probability indicates simply that some neurons are more involved in performing a task than others. Is it hardly surprising that a neuron that responds strongly to one direction of motion and not another is better able to predict performance in a motion discrimination task than a neuron that shows little sensitivity to this distinction.

There is a long history of cognitive models that implement optimal decision making between two-alternatives given noisy evidence (e.g., diffusion models by Ratcliff and colleagues; e.g., Ratcliff, 1978). These models are optimal in the sense that they minimize response times for a given level of accuracy. More recently, neurally inspired models of decision making have been developed that can support optimal performance when the parameters are set optimally (e.g., Shadlen & Newsome, 2001; Usher & McClelland, 2001; Wang, 2002; cf. Bogacz, 2006), and more recently still, some predictions of these models are said to be supported by single-cell recording data. For example, Beck et al. (2008) describe a Bayesian model of decision making in the context of visual motion detection; specifically, deciding whether a set of dots are coherently moving to the left or right when surrounded by noise, in the form of other dots moving randomly. According to their model, neuron in lateral intraparietal (LIP) cortex that preferentially responds to movement in one direction should integrate evidence over time such that (a) the rate of firing grows linearly with time, (b) the rate of growth should be proportional to the strength of the signal (that is, firing rates should change more quickly if a higher proportion of dots are moving in the neurons preferred direction), and (c) the relation between the strength of firing and strength of signal should be maintained when the animal is forced to make a decision in the context of four as opposed to two directions of motion. Beck et al. cite evidence in support of all predictions.

However, there are two general reasons to question how strongly these findings support the Bayesian coding hypothesis. First, contrary to the assumption that humans and monkeys perform optimally in these conditions, there is growing evidence for suboptimal performance under these and similar task conditions. For instance, Zacksenhause, Bogacz, and Holmes (2010) asked human participants to categorize dots moving in one of two directions, and found that only ~30% of the participants performed near optimally. They developed a heuristic process (what they called a robust satisficing solution) that did a better job in accounting for human performance (also see Bogacz, Hu, Holmes & Cohen, 2010; for some related results see Purcell, Heitz, Cohen, Schall, Logan, Palmeri, 2010). Second, it is not clear how diagnostic the neuroscience is in supporting the Bayesian coding hypothesis. The observation that the rate of firing grows more quickly for relevant LIP neurons when the signal is more robust and the fact that this pattern obtains when monkeys need to choose amongst two or four options does not seem too surprising (to us). It is likely that these findings may be explained with alternative (non-Bayesian) theories as well. The finding that LIP neurons show a linear increase in firing as a function of time is perhaps a less obvious prediction, but again, is also predicted by at least one non-Bayesian theory (Grossberg & Pilly, 2008).

Finally, some findings appear to be inconsistent with the Bayesian coding hypothesis. Most notably, Chen, Geisler, and Seidemann (2006) carried out optical imaging of neural firing in V1 of monkeys trained to detect a small visual target. The key finding was that the monkeys' performance was worse than predicted by a Bayesian analysis of the signal recorded in V1, suggesting that the V1 population response is not used optimally by the visual system. Still, Chen et al. (2006) noted two limitations of their analysis that make it difficult to reach any strong conclusions based on their results. On the one hand, the fact that their Bayesian analysis of V1 outperformed the monkeys does not rule out a Bayesian visual system. That is, it is possible the performance of the monkeys was limited by post-visual inefficiencies (perhaps vision is optimal, but motor control is poor). On the other hand, they note the opposite finding (monkeys outperforming the Bayesian analysis) would not rule out a Bayesian visual system either. Although this pattern of results might seem unlikely, the signals recorded by the experimenter inevitably contain only a subset of the information available to the monkey, meaning that a Bayesian analysis on these signals is likely to underestimate ideal performance of the organism. In short, it is difficult to relate neural firing to the Bayesian coding hypothesis. Nevertheless, the Chen et al. (2006) data are perhaps the most relevant evidence to date concerning the Bayesian nature of neural processing within the visual system and suggest that processing past V1 is non-Bayesian (for similar results and conclusions, see Palmer, Cheng, & Seidemann, 2007).

To summarize, there is little evidence from neuroscience that populations of neurons compute in a Bayesian manner. We do not see any reasons to update the previous assessment of Knill, and Pouget (2004) who wrote: “The neurophysiological data on the [Bayesian coding] hypothesis, however, is almost non-existent” (p. 712).

Before concluding this section, we note one implication of the fact that researchers often link Bayesian models in psychology to neuroscience. As noted in Part 1, it is not always clear whether a given researcher is advancing a Methodological or Theoretical Bayesian model. However, the Bayesian coding hypothesis is only relevant to the Theoretical Bayesian perspective. That is, it makes sense to consider the neuroscientific evidence if the claim is that various cognitive, perceptual, and motor systems rely on Bayesian-like algorithms, whereas the neuroscience is irrelevant if Bayesian models are merely a yardstick to measure behavior (with no claim regarding the processes let alone the neural mechanisms that underpin performance). Accordingly, whenever someone highlights the relevance of neuroscience to a Bayesian model, we assume that he or she is endorsing (or at least entertaining) the Theoretical approach.

Part 5: Conceptual problems with the Bayesian approach to studying the mind and brain

As outlined above, there is relatively little evidence in support of Methodological and Theoretical Bayesian theories in psychology and neuroscience: Bayesian theories are so flexible that they can account for almost any pattern of result; Bayesian models are rarely compared to alternative (and simpler) non-Bayesian models and when these approaches are compared non-Bayesian models often do as good a job (or better) in accounting for the data; and the neuroscience data are ambiguous at best. Indeed, the neuroscience is largely irrelevant, given that Bayesian theories in neuroscience are largely inspired by behavioral work in psychology.

But in our view, there is another equally serious problem with both Bayesian approaches that we have only touched on thus far. Bayesian theories in psychology tend to adopt the "rational analysis" methodology of Anderson (1991), according to which the important constraints on cognition come from the nature of the problem and the information available to the organism (the environment). That is, this approach assumes that there is a "triumphant cascade" (Dennett, 1987, p. 227) through Marr's levels of analysis such that the constraints identified at the computational level constitute the main constraints at the algorithmic level. The implication, of course, is that the findings from other domains (e.g., biology and evolution) will play a relatively minor role in constraining theories in psychology. For example, when describing the various steps in carrying out rational analyses, Anderson (1991) writes:

> *The third step is to specify the computational constraints on achieving optimization. This is the true Achilles heel of the rationalist enterprise. It is here that we take into account the constraints that prevent the system from adopting a global optimum. As admitted earlier, these constraints could be complex and arbitrary.* ***To the extent that this is the case, a rationalist theory would fail to achieve its goal, which is to predict behavior from the structure of the environment rather than the structure of the mind****. [bold added] (pp. 473-474)*

In practice, some additional constraints are almost always added to Bayesian models in order that they provide a better (post-hoc) account of human performance, with the constraints often described as general capacity limitations even though specific assumptions are sometimes added (cf. McKenzie, 2003). However, in most cases, these constraints are developed independently of any algorithmic or implementational considerations. In our view, the result is Bayesian models that are not only massively under-constrained but that also mischaracterize cognitive, perceptual, and motor systems[2]. Below we briefly consider constraints from biology and evolution that generally fall outside rational Bayesian considerations.

Constraints from biology

For the sake of illustration, we consider five constraints from biology that should inform psychological theories of vision, the systems most often characterized as optimal. First, the photoreceptors that transduce light energy to neural signals are at the back of the retina, and as a consequence, light has to pass through a number of layers of cells before reaching the photoreceptors, causing shadows and other distortions. This design also results in a blind spot where the optic tract passes through the retina on its way to the thalamus. There is probably some adaptive explanation for this design, but it will be found at the level of biology rather than optics. What is critical for present purposes, however, is that the design of the retina is not only relevant to biologists, but also has important implications for psychological theories of vision. For instance, see Grossberg (1987) for a theory of early vision that is designed, in part, to rectify the distortions of the image that result from the inverted retina.

Second, information projected to the right visual field is projected to the left hemisphere, and vice versa. Indeed, even information projected onto the fovea is split down the middle (cf. Ellis & Brysbaert, 2010). This fact will again provide important constraints for any theory of vision. For starters, it raises the question of how information in the two hemispheres is combined, a form of binding problem. This biological constraint has informed a number of models of visual word identification (e.g., Monaghan & Shillcock, 2008; Shillcock et al., 2000;

[2] For a similar point in the field of behavioural ecology, see McNamara and Houston (2009)

Whitney, 2001), and ultimately all theories of high-level vision will need to confront this constraint. Again, it is hard to see how a rational analysis would constrain an algorithmic theory along these lines.

Third, the visual system is organized into multiple maps, such as retinotopic, ocular dominance, orientation preference, direction of motion, etc. Chklovskii and Koulakov (2004) argue that a wiring optimization economy principle plays a central role in explaining this design. The argument is that it is costly to connect distal neurons with long axons and dendrites, and cortical maps are viewed as solutions that minimize wiring costs. This obviously has important implications for theories of visual processing. For example, once again it raises questions about how the visual system recombines and binds this information. It is not clear that this constraint will arise from a rational analysis.

Fourth, the metabolic cost of firing neurons is high. Lennie (2003) estimated that these costs restrict the brain to activate about 1% of neurons concurrently, and he takes these findings as providing a strong constraint on how knowledge is coded in the cortex. Specifically, Lennie argues that these metabolic costs provide a pressure to learn highly sparse representations of information, consistent with what is found in the cortex (cf. Bowers, 2009). Again, it is not clear how a rational analysis would lead to such a conclusion, particularly given that optimal inference necessitates integrating large amounts of information and not throwing away information.

Fifth, the visual system can identify images of objects and faces in about 100 ms (e.g., Thorpe, Fize, & Marlot, 1996). This provides important constraints on how information flows in the visual system. For example, Thorpe and Imbert (1989) point out that there at least ten synapses separating the retina from object and face selective cells in IT and that neurons at each stage cannot generate more than 1 or 2 spikes in 10 ms. Accordingly, they conclude that neurons at each stage of processing must respond on the basis of one or two spikes from neurons in the previous stage. This in turn limits the amount of feedback that can be involved in identifying an object. These conclusions are unlikely to be reached on the basis of rational analysis.

Of course, there is every reason to assume that biology will be relevant in constraining psychological theories in all domains, and in most cases these constraints will be missed when y considering only the task at hand and the external environment. That is, we suggest that the rationalist approach will miss many of the key constraints required to develop a theory of how the mind works.

Constraints from evolution

Evolution also provides insights into how the mind and brain is organized. Natural selection is the key process by which complex and well designed brains (and other body parts) emerge within a population of organisms. But selection is not constrained to produce optimal solutions. As noted above, Simon (1956) coined the term *satisficing* to highlight the point that natural selection does not produce optimal solutions, but rather solutions that work well enough. As Simon put it:

> *…no one in his right mind will satisfice if he can equally well optimize; no one will settle for good or better if he can have best. But that is not the way the problem usually poses itself in actual design situations (p. 138)*

In a similar way, Dawkins (1982) emphasizes that natural selection is not an optimizing process, and he described the processes as *meliorizing* (from the Latin *melior* meaning ‘better’) rather than satisficing in order to highlight the competitive nature of natural selection. That is, natural selection looks for *better* solutions (better than your competitors), but not the *best* solution.[3]

[3] For three different ways of viewing the relation between evolution and optimality, see Godfrey-Smith (2001).

Critically, a satisficing or meliorizing process is unlikely to produce optimal solutions for two inter-related reasons. First, evolution never starts from scratch, but instead, modifies structures already in place; so-called evolutionary inertia (e.g., Marcus, 2006). Second, evolution is a blind “hill-climbing” process in which minor variants of existing traits are selected that slightly improve the reproductive success of the organism. This hill climbing can, in principle, produce optimal designs when there is only one hill to climb. But when the evolutionary landscape includes multiple peaks and when one starts from the wrong place, then selection will often produce decidedly non-optimal solutions. That is, a hill climbing process combined with evolutionary inertia will often be trapped in a local minimum in an evolutionary landscape. A good example is provided by the three bones that were originally part of the reptilian lower jaw, and used for chewing, but which, in humans, are found in the middle ear, and are used to amplify sounds. As Ramachandran (1990, p. 347) notes:

> *No engineer in his right mind would have come up with such an inelegant solution. The only reason Nature adopted it was because it was available and it worked. It is perhaps not inconceivable that the same sort of thing may have happened time and time again in the evolution of visual mechanisms in the neocortex.*

Not only would no engineer come up with this solution, but neither would a rational analysis constrained only by the environment and the task at hand (see Marcus, 2008, for a range of non-optimal solutions to a wide range of tasks that we face). More generally, we would question a commitment to optimal decision making given that the human phenotype is non-optimal in so many respects. Why would we be (close to) optimal in just this one respect? And even if we were optimal given a poor design, the standard Bayesian approach is incomplete, as it does not provide an explanation for why the design of many systems is so-poor from an engineering point of view. Presumably evolutionary inertia and physiology will provide the key insights here.

When Bayesian models do consider constraints from Biology and Evolution

It should be noted that not all Bayesian models follow the standard logic of rational analysis; that is, the attempt to understand the mind by focusing on the environment and task at hand, with little consideration of other constraint, apart from general processing constraints and specific assumptions inferred from the performance. But from our reading, Bayesian models that are informed by details of biology and evolution are largely restricted to the domain of vision, and for the most part, low-level vision (e.g., Geisler & Diehl, 2003). Interestingly, many researchers in this domain are better described as Methodological rather than Theoretical Bayesians. For instance, when considering Bayesian work directed at pattern detection, discrimination, and identification, Geisler (2009) concludes that "the suboptimal performance of human observers implies substantial contributions of central factors" (p. 3). Indeed, Geisler and colleagues have highlighted in various contexts that perception is often suboptimal after incorporating various constraints from biology (e.g., Chen et al., 2006), and have developed non-Bayesian heuristic models to accommodate behavior that is near optimal (Najemnik, & Geisler, 2009).

Still, Geisler is a strong advocate of the Methodological Bayesian approach because it provides a benchmark for evaluating theories. For instance, Geisler (2011) writes:

> *When human performance approaches ideal performance, then the implications for neural processing can become particularly powerful; specifically, all hypotheses (model observers) that cannot approach ideal performance can be rejected. When human*

> *performance is far below ideal, there are generally a greater number of models than could explain human performance. (p. 772)*

Three points merit consideration here. First, with regards to falsifying theories, the relevance of establishing a Bayesian benchmark is unclear. Consider the case in which human performance is near optimal as determined by a Bayesian model. Under these conditions, a specific algorithmic theory can be rejected because it cannot approach ideal performance as determined by the Bayesian model, or more simply, a model can be rejected because it does not capture human performance. The mismatch in accounting for data is sufficient to rule out the model – the extra step of showing that it does not support optimal performance is superfluous.

Second, we would question the claim that a Bayesian benchmark of near optimal performance provides strong constraints on algorithmic theories that can approximate optimal performance. Consider a recent analysis of perception of speech sounds in noise (Feldman, Griffiths, & Morgan, 2009). The authors are explicit in adopting a rational analysis in which they consider the abstract computational problem, separate from any other algorithmic or biological constraints. They show that an optimal solution results in the reduced discriminability between vowels near prototypical vowel sounds—the so-called perceptual magnet effect that is also observed in humans. At the same time, as noted by the authors themselves, their rational analysis is consistent with algorithmic models that are characterized by feed-forward processing or that include feedback as well as models that include an episodic lexicon or abstract lexicon. That is, the findings are consistent with most types of algorithmic theories that have been advanced independently of any rational analysis. More recently, Shi, Griffiths, Feldman, and Sanborn (2010) showed how a Bayesian model of perception can be implemented with an exemplar model that has a limited vocabulary of items stored in memory. At the same time, they highlight how a previous non-exemplar model by Guenther and Gjaja (1996) predicts highly similar results. In this case, the rational analysis has not strongly constrained the algorithmic level, and we are not aware of any examples in which a Bayesian theory achieves the goal of constraining algorithmic (mechanistic) theories above and beyond the constraints imposed by the data themselves.

Third, as noted by Geisler (2011), the constraints afforded by a Bayesian model get weaker when performance is further from optimal. Low-level vision is presumably one of the better adapted systems, given its long evolutionary history, and accordingly, based on this logic, Bayesian models may be best applied here. However, most other domains, where performance is less likely to be optimal, Bayesian models will be less constraining still.

What are the positive features of Bayesian models?

In our view, the Methodological Bayesian approach does not provide a useful tool for constraining theories, and the Theoretical Bayesian perspective is unsupported. More generally, we claim that it is a mistake to develop models that optimize at a computational level given that the mind/brain satisfices, and given that many of the key constraints for models will be found at the algorithmic and implementational levels. Our main thesis is that Bayesian modeling, both in practice and in principle, is a misguided approach to studying the mind/brain.

Still, it is important to recognize some key contributions of the Bayesian literature. We will briefly highlight three. First, Bayesian modellers have highlighted to the importance of characterizing the environment when developing theories of mind and brain. There is nothing new (or Bayesian) about this. Indeed, this view was taken to the extreme by Gibson (1950) in his “ecological” theory of perception and more recently various (non-Bayesian) theories of statistical learning attempt to exploit information in the world more thoroughly. But we would agree that a failure to fully characterize the environment is common in non-Bayesian theories, potentially leading to fundamental errors. In particular, researchers from a "nativist" perspective may introduce innately specific knowledge if it is (incorrectly) assumed the environment is too impoverished to support learning (cf., Perfors, Tenenbaum, & Regier, 2011).

Second, Bayesian theorists have highlighted the importance of addressing *why* questions. Again, there is nothing new (or Bayesian) about addressing such a fundamental issue, but we would agree that many algorithmic theories in cognitive science have been too focused on accommodating a set of data, with scant attention to the important constraints and insights that can be derived through addressing 'why' questions head on. We think it is unfair of Chater and Oaksford (1999) to characterize non-Bayesian theories as "an assortment of apparently arbitrary mechanisms, subject to equally capricious limitations, with no apparent rationale or purpose" (p. 58), but we do take the point that too many non-Bayesian theories in cognitive science can be characterized as an exercise in curve fitting.

Third, and related to point two above, Bayesian modellers have emphasized "top-down" or "function-first" strategies for theorizing, such that theories are first constrained by the complex computational challenges that organisms face (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). That is, the choice of representations and computations is initially constrained so that it can support core functions in a given domain, with detailed experimental findings used to further constrain a given theory. By contrast, many if not most non-Bayesian models are first constrained by a limited set of data in a given domain, with the goal of elaborating a theory to accommodate more complex phenomena after capturing the more simple phenomena first. This approach might be called a "bottom-up" strategy to theory building.

Again, there is nothing new (or Bayesian) about a top-down strategy per se[4], but we would agree that a failure to give sufficient attention to top-down constraints is a serious limitation of most algorithmic theories in cognition.

To illustrate the contrast between bottom-up and top-down theorizing consider two distinct schools of neural network modelling, namely, the Parallel Distributed Processing (PDP) and symbolic modelling approaches (cf. Holyoak & Hummel, 2000). A key theoretical claim of the PDP approach is that the mind and brain rely on distributed and sub-symbolic representations (cf., Bowers, 2009; Bowers, Damian, & Davis, 2009). These principles allow PDP models to support performance in some restricted domains, such as naming mono-syllable words and nonwords, and this is taken by advocates of this approach as evidence that distributed and sub-symbolic representations support all varieties of cognition, perception, and behavior. That is, an implicit assumption that has guided much of this research program is that it is best to study relatively simple phenomena in great detail and only later attempt to extend any insights to more complex questions (e.g., naming multi-syllable words). This “bottom-up” strategy is common to much (perhaps most) theorizing in psychology.

The obvious danger of this approach, however, it that the representations and processes that support relatively simple tasks may be inadequate to support slightly more complex tasks, such as recognizing familiar words in novel contexts (e.g., CAT in TREECAT; Bowers & Davis, 2009; but see Sibley, Kello, Plaut, & Elman, 2009), let alone more complex language and semantic functions (e.g., Fodor & Pylyshyn, 1988; Pinker & Prince, 1988). Indeed, advocates of “top-down” theorizing have generally argued that more complex language functions require

[4] What is relatively unique about the Bayesian approach to “top-down” theorizing is that it is carried out a computational level rather than an algorithmic level. Still, at whatever level a theory is developed, the key point is that complex (and functional) task demands are used as an initial constraint for theory building.

localist and context independent (symbolic) representations, and based on these functional constraints, have developed “localist” and “symbolic” networks (e.g., Davis, 2010; Grossberg, 1980; Hummel & Holyoak, 1997). Of course, the role of local and symbolic representations in cognition is far from settled, but it is interesting to note that the position one takes on the local vs. distributed and symbolic vs. non-symbolic debates closely follows whether the theorist adopts a “top-down” or “bottom-up” strategy of theorizing. Perhaps it is no coincidence that Bayesian models designed to address more complex functional problems often include structured (componential) representations in their computational theories (Kemp & Tenenbaum, 2009).

In sum, we are sympathetic with many of the research strategies emphasized by advocates of Bayesian modeling. But if the “Bayesian revolution” (Shultz, 2007, p. 357) is thought to provide a new method or a new theory of mind, it has to be something above and beyond these familiar (albeit too often ignored) research practices.

Conclusions

A broad range of evidence has been marshaled in support of the claim that the mind and brain perform Bayesian statistical reasoning (or something similar) in order to make or approximate optimal decisions given noisy data. Yet in our view, the evidence is rather weak. Clearly, though, proponents of the Bayesian approach have reached quite different conclusions on the basis of the same evidence. To conclude, we consider some factors that might explain these discrepant views. To do so, we make use of Bayes’s rule, which provides a rational way to evaluate the probability of the hypothesis that humans are optimal Bayesian estimators, given the available evidence. Our goal, of course, is not to derive a specific probability estimate for this hypothesis. Rather, we adopt this approach because it is helpful in highlighting where we think there are problems in the way Bayesian theorists have evaluated the evidence, and why these problems lead to a greatly exaggerated assessment of the probability of the hypothesis.

The argument is sketched in Figure 2, which depicts Bayes's rule. The hypothesis under consideration is that humans are optimal (or near-optimal) Bayesian observers; we shall refer to this hypothesis as $H_{optimal}$. We seek to evaluate $P\ (H_{optimal} \mid E)$, the posterior probability of this hypothesis given the evidence (E). Our contention is that there are three problems with the way Bayesian theorists have evaluated this posterior probability. These three problems are directly related to the three components of the right-hand side of the equation in Figure 2. The first two problems overestimate the numerator of the equation, while the third problem underestimates the denominator. We consider these problems in turn.

First, Bayesian theorists have overestimated the prior probability $P\ (H_{optimal})$. This prior probability is not explicitly considered by Bayesian theorists and seems, implicitly, to be accorded the same probability as the alternative hypothesis $P\ (\sim H_{optimal})$ (i.e., that humans are not optimal Bayesian estimators). We argue that a rational choice of prior for $H_{optimal}$ should be very low, based on previous evidence concerning the flaws in human reasoning and perception, as well as the biological, evolutionary, and behavioral constraints considered in Part 5. Indeed, it is the surprising nature of the claim—that we are close to optimal—that has contributed to the great interest in Bayesian theories. This surprise presumably reflects the low-prior that most people give to this hypothesis.

Second, the likelihood $P(E|H_{optimal})$ is typically overestimated relative to the likelihood that would be computed on the basis of purely rational considerations. Overestimation of the likelihood is possible because of the lack of constraint in Bayesian theories. Indeed, as detailed in Part 2, priors, likelihoods, etc. are often selected post-hoc in order to account for the data themselves. In effect, there is a family of $H_{optimal}$ hypotheses, corresponding to the different choices of priors, likelihood functions, and so on, and modellers are at liberty to choose the specific $H_{optimal}$ that maximizes the likelihood of the evidence given the hypothesis. A related problem noted in Part 2 is the tendency to explain failures of the optimal model by reference to auxiliary assumptions or to effects at the algorithmic level. In such cases the likelihood $P(E|H_{optimal})$ is artificially maintained. These problems in the estimation of the likelihood of the evidence undermine the claim that human behavior approximates, in any meaningful sense, optimal performance.

The third problem concerns the underestimation of $P(E)$, the marginal probability of the evidence. This term can be expanded as $P(E) = P(H_{optimal})$ x $P(E|H_{optimal}) + P(\sim H_{optimal})$ x $P(E|\sim H_{optimal})$. Bayesian theorists have focused on the first part of this expansion, but have neglected the second joint probability, which expresses the probability that the same evidence could be explained by alternative, non-Bayesian hypotheses. As shown in Part 3, there is little evidence that Bayesian theories provide a better fit to the data than various alternative non-Bayesian accounts. Accordingly, even if a Bayesian model can account for human performance in a principled way, the conclusion that human performance is mediated by Bayesian processes is unjustified.

Together, these considerations undermine the strong claims often made by Methodological and Theoretical Bayesian theorists, claims such as those highlighted in quotes at the start of this paper. The consequence of overestimating the prior probability and likelihood of $H_{optimal}$ and underestimating the probability of the evidence by ignoring alternative explanations is that Bayesian theorists greatly overestimate the posterior probability of the hypothesis that humans are optimal (or near optimal) Bayesian estimators. The result, in our view, is a collection of Bayesian “just-so” stories in psychology and neuroscience; that is, mathematical analyses of cognition that can be used to explain almost any behavior as optimal. If the data had turned out otherwise, a different Bayesian model would have been constructed to justify the same conclusion that human performance is close to optimal.

More generally, it is not clear to us how Methodological or Theoretical Bayesian approaches can provide additional insights into the nature of mind and brain compared to non-Bayesian models constrained by adaptive (rather than optimality) considerations. Both Bayesian and non-Bayesian theories can offer answers to ‘why’ questions, and the relevant metric to evaluate the success of a theory is actual behavior, not optimal performance by some criterion. Theories in psychology and neuroscience should be developed in the context of all variety of constraints, including those afforded by the algorithmic and implementational levels of description. In our view, the Bayesian approach of focusing on optimality at a computational level of description leads to under-constrained theories that often mischaracterize the systems that support cognition, perception, and behavior.

References

Amarasingham, A., Chen, T. L., Geman, S., Harrison, M. T., & Sheinberg, D. L. (2006). Spike count reliability and the Poisson hypothesis. *Journal of Neuroscience, 26*(3), 801-809. doi:10.1523/JNEUROSCI.2948-05.2006

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

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the Environment in Memory. *Psychological Science, 2*(6), 396-408.

Balota, D.A., Neely H.H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6*, 576–587. doi: 10.1037/0278-7393.6.5.576

Bar-Eli, M., Azar, O. H., & Lurie, Y. (2009). (Ir)rationality in action: do soccer players and goalkeepers fail to learn how to best perform during a penalty kick? *Progress in Brain Researcj, 174*, 97-108.doi:10.1016/S0079-6123(09)01309-0

Bartlett, F.C.(1932*). Remembering: A study in experimental and social psychology*. Cambridge University Press, Cambridge.

Becker, G. (1991). The nonoptimality of Anderson's memory fits. Commentary on Anderson (1991) " Is human cognition adaptive". *Behavioral and Brain Sciences*, 14, 487-488.

Bogacz, R. (2006). Optimal decision-making theories: linking neurobiology with behavior. *Trends in Cognitive Sciences*, *11*, 118–125. doi:10.1016/j.tics.2006.12.006

Bogacz, R., Hu, P., Holmes, P., & Cohen, J. (2010). Do humans produce the speed-accuracy tradeoff that maximizes reward rate? *Quarterly Journal of Experimental Psychology*, *63*, 863–891. doi: 10.1080/17470210903091643

Bowers, J. S. (2009). On the Biological Plausibility of Grandmother Cells: Implications for Neural Network Theories in Psychology and Neuroscience. *Psychological Review, 116*(1), 220-251. doi: 10.1037/a0014462

Bowers, J. S. (2010a). Does masked and unmasked priming reflect Bayesian inference as implemented in the Bayesian Reader? *European Journal of Cognitive Psychology*, *22*, 779-797. doi:10.1080/09541446.2010.532120

Bowers, J.S. (2010b). What are the Bayesian constraints in the Bayesian reader? Reply to Norris and Kinoshita (2010). *European Journal of Cognitive Psychology*, *22*, 1270-1273. doi:10.1080/09541446.2010.532120

Bowers, J.S., Damian, M.F., & Davis, C.J. (2009). A fundamental limitation of the conjunctive codes learned in PDP models of cognition: Comments on Botvinick and Plaut (2006). *Psychological Review, 116*, 986-995. doi: 10.1037/a0017097

Bowers, J.S., & Davis, C.J. (2009) Learning representations of wordforms with recurrent networks: Comment on Sibley, Kello, Plaut, & Elman (2008). *Cognitive Science, 33,* 1183-1186. doi: 10.1111/j.1551-6709.2009.01062.x

Bowers, J.S. & Davis, C.J. (2011). More varieties of Bayesian theories, but no enlightenment. Commentary on "Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition". *Behavioral and Brain Sciences, 34*, 193-194. doi:10.1017/S0140525X11000227

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Automatic semantic activation of embedded words: Is there a ‘hat’ in ‘that’? *Journal of Memory and Language*, *52*, 131–143. http://dx.doi.org/10.1016/j.jml.2004.09.003

Bowers, J. S., Mattys, S. L., & Gage, S. H. (2009). Preserved implicit knowledge of a forgotten childhood language. *Psychological Science, 20*, 1064-1069. doi: 10.1111/j.1467-9280.2009.02407.x

Brighton, H,. & Gigerenzer, G. (2008). Bayesian brains and cognitive mechanisms: harmony or dissonance? In: N. Chater & M. Oaksford (Eds.) *The probabilistic mind: prospects for Bayesian cognitive science*. Oxford University Press, New York, pp 189–208.

Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., & Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Visual Neuroscience, 13, 87–100.

Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, *12*, 65-86. doi: 10.1080/095414400382208

Burr, D., Banks, M.S., & Morrone, M.C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, *198*, 49-57. doi: 10.1007/s00221-009-1933-z

Chater, N. (2000). How smart can simple heuristics be? Commentary on Todd, P.M, & Gigerenzer G. (2000). Précis of Simple heuristics that make us smart. *Behavioral and Brain Sciences, 23*, 745-746.

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

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

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: where next? *Trends in Cognitive Sciences, 10*(7), 292-293. doi: 10.1016/j.tics.2006.05.008

Chen, Y. Z., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neuroscience, 9,* 1412-1420. doi:10.1038/nn1792

Chklovskii, D. B., & Koulakov, A. A. (2004). Maps in the brain: What can we learn from them? *Annual Review of Neuroscience, 27*, 369-392. doi: 10.1146/annurev.neuro.27.070203.144226

Davis, C. J. (2006). Orthographic input coding: A review of behavioural data and current models. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 180–206). New York: Psychology Press.

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

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception & Performance, 32,* 668-687. doi: 10.1037/0096-1523.32.3.668

Dawkins, R. (1982). *The extended phenotype: The gene as the unit of selection*. Oxford; San Francisco: W.H. Freeman.

Dennett, D. (1987): *The International Stance*. Cambridge, MA, MIT Press.

DeWeese, M. R., Wehr, M., & Zador, A. M. (2003). Binary spiking in auditory cortex. *Journal of Neuroscience, 23*(21), 7940-7949.

Eddy, D.M. (1982). Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities. In D. Kahneman, P. Slovic and A. Tversky (Eds), *Judgement under Uncertainty: Heuristics and Biases*. New York, Cambridge University Press.

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 1103–1123. doi:10.1037/0278-7393.26.5.1103

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

Evans, J.St.B.T, & Over, D (2004). *If.* Oxford, England: Oxford University Press.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9*(4), 292-303. doi: 10.1038/nrn2258

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

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition, 109,* 416–422. doi:10.1016/j.cognition.2008.09.007

Geisler, W.S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51,* 771-781. doi:10.1016/j.visres.2010.09.027

Geisler, W. S., & Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, *27*(3), 379–402. http://dx.doi.org/10.1016/S0364-0213(03)00009-0

Geisler, W.S. & Ringach, D. (2009). Natural Systems Analysis. *Visual Neuroscience*, *26*, 1-3. doi: 10.1017/S0952523808081005

Gibson, J.J.,(1950). *The Perception of the Visual World*. Riverside Press, Cambridge.

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science, 1*,1, 107-143. doi: 10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8, 53-96. doi: 10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., & Selten, R. (2001). Rethinking rationality. In G. Gigerenzer & R. Selten (Eds.), *Bounded Rationality - the Adaptive Tool* (pp. 1-12). Cambridge: M I T Press.

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

Godfrey-Smith, P. (2001). Three Kinds of Adaptationism, in S. H. Orzack and E. Sober (eds.), *Adaptationism and Optimality*. Cambridge: Cambridge University Press.

Gold, J.M., Tadin, D., Cook, S.C., &. Blake, R.B. (2008). The efficiency of biological motion perception. *Attention, Perception & Psychophysics, 70*, 88–95. doi: 10.3758/PP.70.1.88

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*(1), 75-90. doi: 10.1037/0033-295X.109.1.75

Grainger, J., Granier, J. P., Farioli, F., Van Assche, E., & van Heuven, W. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 865–884 doi:10.1037/0096-1523.32.4.865

Grossberg, S. (1980). How does the brain build a cognitive code? *Psychological Review, 87*, 1–51. doi: 10.1037/0033-295X.87.1.1

Grossberg, S. (1987). Competitive learning—from interactive activation to adaptive resonance. *Cognitive Science, 11*, 23–63. doi: 10.1111/j.1551-6708.1987.tb00862.x

Grossberg, S. and Pilly, P. (2008). Temporal dynamics of decision-making during motion perception in the visual cortex. *Vision Research, 48*, 1345-1373. doi:10.1016/j.visres.2008.02.019

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America, 100*, 1111-1121. doi: 10.1121/1.416296

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. http://dx.doi.org/10.1016/j.tics.2010.05.004

Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008). Bayesian models of cognition. In Ron Sun (ed.), *Cambridge Handbook of Computational Cognitive Modeling*. Cambridge University Press.

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

Grossberg, S. (1987). Cortical Dynamics Of 3-Dimensional Form, Color, and Brightness Perception .1. Monocular Theory. *Attention, Perception & Psychophysics, 41*, 87-116. doi: 10.3758/BF03204874

Gu, Y., DeAngelis, G.C., Angelaki, D.E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nature Neuroscience, 10*, 1038-1047. doi:10.1038/nn1935

Gur, M., Beylin, A., & Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (V1) of alert monkeys. *Journal of Neuroscience, 17*(8), 2914-2920.

Gur, M., & Snodderly, D.M. (2006). High response reliability of neurons in primary visual cortex (V1) of alert, trained monkeys. *Cerebral Cortex, 16*, 888-895. doi: 10.1093/cercor/bhj032

Hammett, S. T., Champion, R. A., Thompson, P. G., & Morland, A. B. (2007). Perceptual distortions of speed at low luminance: Evidence inconsistent with a Bayesian account of speed encoding. *Vision Research, 47*, 564–568. doi:10.1016/j.visres.2006.08.013

Hahn, U. & Warren, P.A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review*. *116*, 454-461. doi: 10.1037/a0015241

Holyoak, K. J., & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Deitrich & A. Markman (Eds.), Cognitive dynamics: Conceptual change in humans and machines (pp. 229–263). Mahwah, NJ: Erlbaum

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427-466. doi: 10.1037/0033-295X.104.3.427

Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review, 16*, 80-87. doi: 10.3758/PBR.16.1.80

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

Jazayeri, M. (2008) Probabilistic sensory recoding. *Current Opinion in Neurobiology, 18,* 431–437. Doi: 10.1016/j.conb.2008.09.00

Johnson, J. & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99. doi:10.1016/0010-0285(89)90003-0

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*, 193-194. doi:10.1017/S0140525X10003134

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. Psychological Review, 116 (4), 856-874. doi:10.1037/a0016979

Juhasz, B.J., & Rayner, K (2006). The role of age-of-acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition, 13,* 846-863. doi: 10.1080/13506280544000075

Kara, P., Reinagel, P. & Reid, R. C. (2000). Low response variability in simultaneously recorded retinal, thalamic and cortical neurons. *Neuron, 27,* 635–646. doi:10.1016/S0896-6273(00)00072-6

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics.. *American Economic Review* 93,1449-14475.

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

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591. doi: 10.1037/0033-295X.103.3.582

Kemp, C. and Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. Psychological Review, 116, 20-58. doi: 10.1037/a0014282

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

Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, *113,* 123–127. doi:10.1016/j.cognition.2009.07.00

Kording, K. P., & Wolpert, D. M. (2004). The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences of the United States of America, 101,* 9839-9842. doi:10.1073/pnas.0308394101

Kording, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10,* 1364-6613. . doi:10.1016/j.tics.2006.05.003

Lee, M.D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review, 15,* 1-15. doi: 10.3758/PBR.15.1.1

Lennie, P. (2003). The cost of cortical computation. *Current Biology, 13*, 493-497. http://dx.doi.org/10.1016/S0960-9822(03)00135-0.

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*(6), 969-998. doi: 10.1111/j.1551-6709.2009.01045.x

Lieberman, D., Tooby, J. & Cosmides, L. (2007). The architecture of human kin detection. *Nature, 445,* 727-731. doi:10.1038/nature05510

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*(11), 1432-1438. doi:10.1038/nn1790

Ma, W. J., & Pouget, A. (2008). Linking neurons to behavior in multisensory perception: A computational review. *Brain Research, 1242,* 4-12. doi:10.1016/j.brainres.2008.04.082

Maimon, G., & Assad, J. A. (2009). Beyond Poisson: Increased Spike-Time Regularity across Primate Parietal Cortex. *Neuron, 62,* 426-440. doi:10.1016/j.neuron.2009.03.021

Mamassian, P. (2008). Overconfidence in an objective anticipatory motor task. *Psychological Science, 19,* 601-606. doi: 10.1111/j.1467-9280.2008.02129.x

Marcus, G. F. (2006). Cognitive Architecture and Descent with Modification. *Cognition* 101, 443-465. doi:10.1016/j.cognition.2006.04.009

Marcus, G. (2008). *Kluge: The haphazard construction of the mind.* Boston: Houghton Mifflin Company.

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

Marteau, T. M. (1989). Framing of Information - Its Influence Upon Decisions of Doctors and Patients. *British Journal of Social Psychology, 28,* 89-94. doi: 10.1111/j.2044-8309.1989.tb00849.x

Maloney, L.T. and Mamassian, P. (2009) Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. Visual Neuroscience, 26, 147–155. doi:10.1017/S0952523808080905

Masse, N.Y., & Cook, E. P. (2008). The effect of middle temporal spike phase on sensory encoding and correlates with behavior during a motion-detection task. Journal of Neuroscience, 28, 1343–1355. doi:10.1523/JNEUROSCI.2775-07.2008

McClelland, J. L. (2006). How far can you go with Hebbian learning, and when does it lead you astray? In Munakata, Y. & Johnson, M. H. Processes of Change in Brain and Cognitive Development: *Attention and Performance XXI*. pp. 33-69. Oxford: Oxford University Press.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457. doi: 10.1037/0033-295X.102.3.419

Monaghan, P. & Shillcock, R. (2008). Hemispheric dissociation and dyslexia in a computational model of reading. *Brain and Language, 107*, 185–193. doi:10.1016/j.bandl.2007.12.005

Morrison, C. M., & Ellis, A. W. (1995). The roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 116-133. doi: 10.1037/0278-7393.21.1.116

Movellan, J. R., & Nelson, J. D. (2001). Probabilistic functionalism: A unifying paradigm for the cognitive sciences. *Behavioral and Brain Sciences, 24,* 690-691.

Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal Predictions in Everyday Cognition: The Wisdom of Individuals or Crowds? *Cognitive Science, 32*(7), 1133-1147. doi: 10.1080/03640210802353016

Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search*. Vision Research, 49,* 1286-1294. doi:10.1016/j.visres.2008.12.005

Nelson, J.D. (2009). Naïve optimality: Subjects' heuristics can be better-motivated than experimenters' optimal models. *Behavioral and Brain Sciences, 32,* 94-95.

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

Norris, D. & Kinoshita, S. (2007). Slow slots with slops: Evidence for 'slot-coding' of letter positions with positional noise. *Proceedings of the European Society for Cognitive Psychology*, Marseille, August, 2007.

Norris, D., Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General. 137(3),* 434-455. doi: 10.1037/a0012799

Norris, D., & Kinoshita, S. (2010). Explanation versus accommodation: Reply to Bowers (2010). *European Journal of Cognitive Psychology*, *22*, 1261-1269. doi: 10.1080/09541446.2010.524201

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 117,* 357–395. doi: 10.1037/0033-295X.115.2.357

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

Oaksford, M. & Chater, N. (2003). Computational levels and conditional inference: Reply to Schroyens and Schaeken (2003). *Journal of Experimental Psychology: Learning, Memory & Cognition, 29*, 150-156, doi: 10.1037/0278-7393.29.1.150

Oaksford, M. & Chater, N. (2007). *Bayesian Rationality*. Oxford University Press: Oxford.

Palmer, C.R., Cheng, S. Y., & Seidemann, E. (2007). Linking neuronal and behavioral performance in a reaction-time visual detection task. *The Journal of Neuroscience*, *27* (30), 8122-8137. doi: 10.1523/JNEUROSCI.1940-07.2007

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature, 423,* 752-756. doi:10.1038/nature01516

Perfors, A, Tenenbaum, J. B. and Regier, T. (2011). Cognition.The learnability of abstract syntactic principles. Cognition 118,306–333. doi:10.1016/j.cognition.2010.11.001

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language-acquisition. *Cognition*, *28*, 73–193. doi:10.1016/0010-0277(88)90032-7

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472– 491. doi: 10.1037/0033-295X.109.3.472

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*, 1113–1143. doi: 10.1037/a0020311

Purushothaman G, & Bradley, D.C. (2005). Neural population code for fine perceptual decisions in area MT. *Nature Neuroscience, 8,* 99-106. doi:10.1038/nn1373

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

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367. doi: 10.1037/0033-295X.107.2.358

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review, 12,* 573–604. doi: 10.3758/BF03196750

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*, 734-760. doi: 10.1037/0033-295X.103.4.734

Seydell, A., Knill, D. C., & Trommershauser, J. (2010). Adapting internal statistical models for interpreting visual cues to depth. *Journal of Vision, 10*(4). doi: 10.1167/10.4.1

Schacter, D.L. (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston: Houghton Mifflin.

Scherag, A., Demuth, L. Roesler, F., Neville, H. J. & Roeder, B. (2004) The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition 93*, B97-B108. doi:10.1016/j.cognition.2004.02.003

Shackleton, T. M., Skottun, B. C., Arnott, R. H., & Palmer, A. R. (2003). Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of guinea pigs. *Journal of Neuroscience, 23,* 716-724.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience, 18,* 3870-3896.

Shadlen, M.N., & Newsome, W.T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology, 86*, 1916–193.

Shiffrin, R.M. & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving Effectively from Memory. *Psychonomic Bulletin & Review, 4*, 145-166. doi: 10.3758/BF03209391

Shillcock, R., Ellison, T.M., & Monaghan, P. (2000). Eye-fixation behavior, lexical storage and visual word recognition in a split processing model. *Psychological Review, 107,* 824–851. doi: 10.1037/0033-295X.107.4.824

Shultz, T.R. (2007). The Bayesian revolution approaches psychological development. *Developmental Science, 10),* 357–364. doi: 10.1111/j.1467-7687.2007.00588.x

Sibley, D.E., Kello, C. T., Plaut, D. C., & Elman, J. L (2009). Sequence encoders enable large-scale lexical modeling: Reply to Bowers and Davis (2009). *Cognitive Science*, *33*, 1187-1191. doi: 10.1111/j.1551-6709.2009.01064.x

Sikstrom, S. (1999). Power function forgetting curves as an emergent property of biologically plausible neural network models. *International Journal of Psychology, 34*, 460–464. doi: 10.1080/002075999399828

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129–138. doi: 10.1037/h0042769

Sloman, S. A. & Fernbach, P. M. (2008). The value of rational analysis: An assessment of causal reasoning and learning. In Chater, N. & Oaksford, M. (Eds.). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford: Oxford University Press.

Stadthagen-Gonzalez, H., Bowers, J. S., & Damian, M. F. (2004). Age-of-acquisition effects in visual word recognition: evidence from expert vocabularies. *Cognition, 93*(1), B11-B26. doi:10.1016/j.cognition.2003.10.009

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. (2002). Communicating accuracy of tests to general practitioners: a controlled study. *British Medical Journal, 324*, 824–826. doi: 10.1136/bmj.324.7341.824

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*, 578-585. doi:10.1038/nn1669

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology, 53,* 168-179. doi:10.1016/j.jmp.2008.11.002

Thompson, P., Brooks, K., & Hammett, S. T. (2006). Speed can go up as well as down at low contrast: implications for models of motion perception. *Vision Research*, *46*, 782–786. doi:10.1016/j.visres.2005.08.005

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522. doi: 10.1038/381520a0

Thorpe S, Imbert M. (1989). Biological constraints on connectionist modelling. In R. Pfeifer, F. Fogeiman-Soule, S. Steels, & Z. Schreter (Eds.) *Connectionism in* *perspective*. Amsterdam: Elsevier/North-Holland, 63-93.

Trommershauser, J., Maloney, L. T., & Landy, M. S. (2003). The consistency of bisection judgments in visual grasp space. *Journal of Vision, 3*, 795-807. doi: 10.1167/3.11.13

Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science. 211*, 453-458. doi: 10.1126/science.7455683

Usher, M., & McClelland, J.L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review, 108*, 550–592. doi: 10.1037/0033-295X.108.3.550

Wagenmakers, E.-J. (2009). How do individuals reason in the Wason card selection task? Comment on "Bayesian rationality: The probabilistic approach to human reasoning". Behavioral and Brain Sciences, 32, 104.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology, 60*, 158-189. doi:10.1016/j.cogpsych.2009.12.001

Wang, X.J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron, 36*, 955–968. doi:10.1016/S0896-6273(02)01092-9 |

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*, 598-604. doi: 10.1038/nn85

West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition, 31*, 243–51. doi: 10.3758/BF03194383

Whitney, C. (2001) How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review, 8*, 221-243. doi: 10.3758/BF03196158

Wilson, D. K., Kaplan, R. M., & Schneiderman, L. J. (1987). Framing of Decisions and Selections of Alternatives in Health-Care. *Social Behaviour, 2*, 51-59.

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science, 18*(2), 133-134. doi: 10.1111/j.1467-9280.2007.01862.x

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science, 2*(6), 409-415. doi: 10.1111/j.1467-9280.1991.tb00175.x

Wolpert, D. M. (2007). Probabilistic models in human sensorimotor control. *Human Movement Science, 26*, 511-524. doi:10.1016/j.humov.2007.05.005

Wu, S. W., Trommershauser, J., Maloney, L. T., & Landy, M. S. (2006). Limits to human movement planning in tasks with asymmetric gain landscapes. *Journal of Vision, 6*(1), 53-63. doi: 10.1167/6.1.5

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

Figure Captions

Figure 1: Separate probability distributions are shown for the prior, the posterior and the likelihood function. The posterior probability effectively revises the prior probability in the direction of the likelihood function.

Figure 2: A Bayesian analysis of the problems with the way Bayesian computation models have used evidence to evaluate the hypothesis that humans are optimal Bayesian estimators.

Figure 1

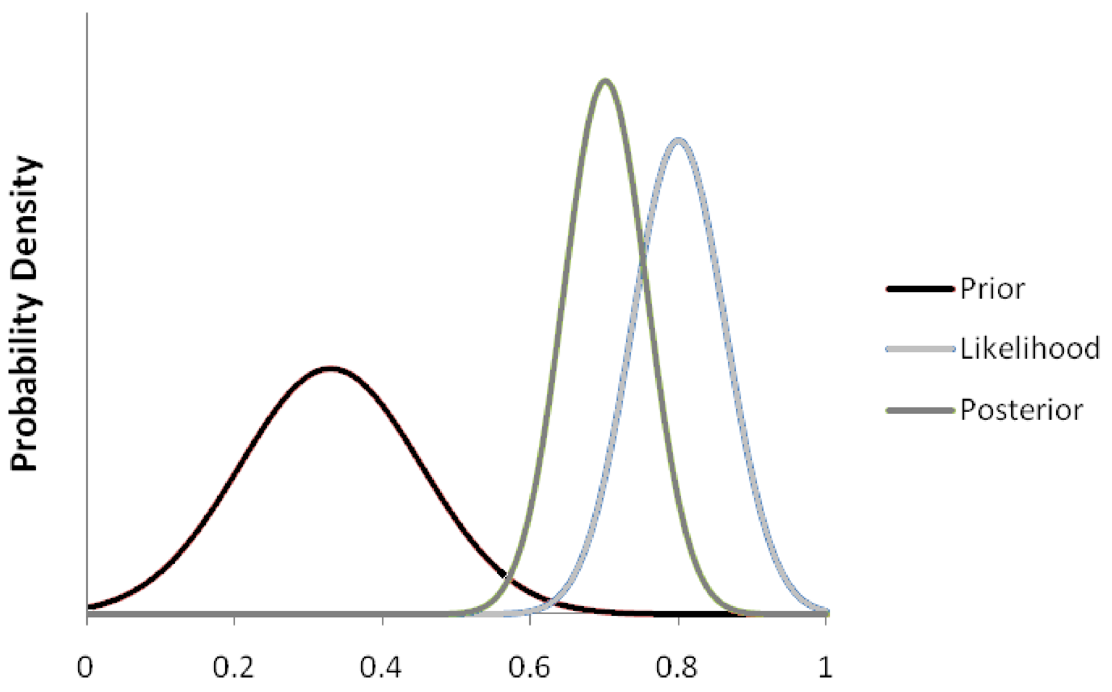
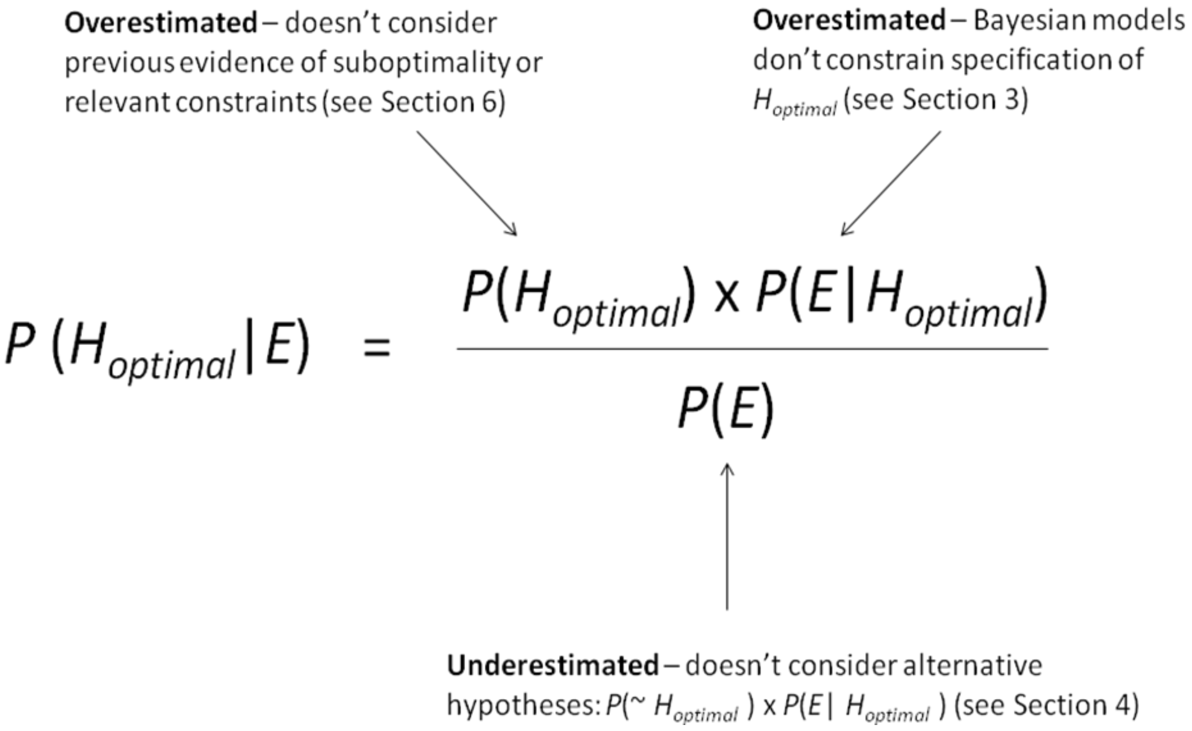

Probability Density
0
0.2
0.4
0.6
0.8
1
Prior
Likelihood
Posterior

Figure 2



Overestimated – doesn’t consider previous evidence of suboptimality or relevant constraints (see Section 6)

Overestimated – Bayesian models don’t constrain specification of $H_{optimal}$ (see Section 3)

$$P\,(H_{optimal}|E) \quad = \quad \frac{P(H_{optimal}) \text{ x } P(E|H_{optimal})}{P(E)}$$

Underestimated – doesn’t consider alternative hypotheses: $P(\sim H_{optimal})$ x $P(E|\ H_{optimal})$ (see Section 4)