Psychological Review
2005, Vol. 112, No. 3, 662–668

0033-295X/05/$12.00 DOI: 10.1037/0033-295X.112.3.662

Bayesian Statistical Inference in Psychology: Comment on Trafimow (2003)

Michael D. Lee
University of Adelaide

Eric-Jan Wagenmakers
University of Amsterdam

D. Trafimow (2003) presented an analysis of null hypothesis significance testing (NHST) using Bayes's theorem. Among other points, he concluded that NHST is logically invalid, but that logically valid Bayesian analyses are often not possible. The latter conclusion reflects a fundamental misunderstanding of the nature of Bayesian inference. This view needs correction, because Bayesian methods have an important role to play in many psychological problems where standard techniques are inadequate. This comment, with the help of a simple example, explains the usefulness of Bayesian inference for psychology.

Keywords: Bayesian statistics, Bayes factor, frequentist statistics, null hypothesis significance testing, transformational invariance

Like all empirical sciences, psychology deals with uncertainty. Samples are limited, measurements are imprecise, and confounding variables are difficult to control. Indeed, each of these sources of uncertainty can be particularly severe under the experimental and ethical constraints involved in studying psychological phenomena. This means psychology relies heavily on statistical inference, and its success as a discipline depends crucially on the quality of the methods it uses.

In this context, Trafimow (2003) considered the relationship between standard statistical practice and Bayesian approaches. A central part of Trafimow's conclusion is that null hypothesis significance testing (NHST) is logically invalid but that logically valid Bayesian analyses are often not possible. In particular, he argued, "My best guess is that some of the necessary information, particularly $p(H_0)$ and $p(F\backslash\text{–}H_0)$ or numbers from which these probabilities can be estimated, are often lacking, and consequently a Bayesian approach cannot be used" (p. 534), and thus "it is understandable that NHSTP [NHST procedure] rather than Bayesian analysis is the order of the day—after all, researchers can perform the analyses" (p. 527).

It is difficult to imagine how two sentences could more fundamentally misrepresent Bayesian statistical inference. Bayesian analysis is not simply a way of shuffling the probabilities used in standard statistics, together with some new probability statements representing things like priors, in accordance with Bayes's theorem. In fact, Bayesian statistical inference is philosophically and fundamentally different from the standard approach. Its differences allow Bayesian inference to address problems standard approaches cannot, and make it more general and efficient than the canned recipes that dominate current practice in psychology. Most importantly, and contrary to Trafimow's (2003) assertion, the basic nature of Bayesian inference guarantees that it is always possible to undertake analyses.

Trafimow's (2003) mistaken claims need correction, because psychology now has much to benefit from the adoption of Bayesian methods. With the advent of modern computing technology, Bayesian inference has become possible in practice as well as in principle and has been adopted with great success in most scientific fields. The literature now contains literally thousands of worked examples—ranging from astrophysics (e.g., Gregory, 1999) to particle physics (e.g., Sivia, Carlile, & Howells, 1992) to image processing (e.g., Gull & Daniell, 1980) to economics (e.g., Zellner, 1984)—where Bayesian techniques have successfully been used to tackle problems that were previously difficult or impossible using standard methods. In psychology, Bayesian methods have recently been successfully applied to the fundamental problem of choosing between competing hypotheses or models of psychological data (e.g., Myung, Forster, & Browne, 2000; Myung & Pitt, 1997; Pitt, Myung, & Zhang, 2002).

In this comment, we describe, with a simple example, those situations in psychological research where Bayesian methods are more useful and applicable than standard methods. We emphasize that Bayesian analyses are always possible in principle, and that current computational methods make most Bayesian analyses feasible for the sorts of problems psychology addresses.

The Strengths of Bayesian Inference

The strength of standard statistical approaches lies in their computational simplicity, which stems from having been developed before the advent of modern computing abilities. Standard methods consist of a set of usually sensible, but always ad hoc, heuristics for estimating uncertain quantities, and testing the significance of observations by comparing them with those that would occur by chance.

Michael D. Lee, Department of Psychology, University of Adelaide, Adelaide, Australia; Eric-Jan Wagenmakers, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands.

We thank Geoffrey Loftus, Dan Navarro, Mark Pitt, and David Trafimow for helpful comments.

Correspondence concerning this article should be addressed to Michael D. Lee, Department of Psychology, University of Adelaide, Adelaide, South Australia 5005, Australia. E-mail: michael.lee@adelaide.edu.au

Jaynes (2003, p. 550) argued that standard methods are usable and useful when dealing with the problems for which they were originally developed. These are basically problems where all the relevant information comes (or can accurately be conceived of as coming) from independent runs of simple random experiments. This means that standard methods work well when three conditions are met:

1. The variables of interest vary according to simple distributions, like the normal and binomial, in which a small number of parameters adequately describe their distributional form.

2. No important prior information is available about the variables of interest.

3. The number of data is reasonably large.

When these conditions are met, standard methods are often quick, easy, and useful. Outside of these conditions, they behave like most heuristics when applied to unintended problems. At best, they are inefficient. At worst, they are inapplicable or produce pathological behavior that makes them useless or misleading. The literature now contains many concrete examples of inefficiencies and pathologies for both estimators and significance tests for realistic problems of scientific inference (e.g., Jaynes, 2003; Lindley, 1972; Lindley & Phillips, 1976; Press, 1989).

The key point, of course, is that psychology has many important problems that do not meet one or more of these three conditions:

1. It seems very unlikely that every interesting psychological variable has a simple distributional form. There is considerable debate in the literature, for example, about the distributional form of response times (e.g., Van Zandt, 2000).

2. Because psychology is inherently interdisciplinary, relevant prior information often comes from other fields. For example, basic chemical knowledge must be relevant for many drug-related studies in psychology. Sometimes this information takes the form of a lawful relationship, rather than data coming from an experiment. Unlike standard methods, Bayesian statistics deals naturally with both types of information.

3. In clinical and other applied settings, it is sometimes difficult, expensive, or just impossible to obtain data in any great number.

Unlike standard methods, Bayesian inference has a principled basis in probability theory, as formalized by Cox's (1961) theorems. This means it can be applied to any statistical distribution, naturally incorporates relevant prior information, and is valid for any number of data. Bayesian methods, therefore, have the potential to enhance and further psychological understanding, allowing new and different problems to be tackled.

The Process of Bayesian Inference

Bayesian inference explicitly represents what is known and unknown about variables of interest, at all stages of analysis, using probability distributions. For variables that take discrete values, these distributions assign a probability value to each possibility. For variables that can assume a continuous range of values, the distributions assign a probability density value to each possibility. Probability densities behave much like probabilities: Ratios of densities quantify how much more likely one outcome is than another, and sums of densities over a range of outcomes give the probability that the variable lies in that range. In both the discrete and continuous cases, probability distributions naturally and intuitively represent the current knowledge and remaining uncertainty researchers have about psychological variables of interest.

Bayesian analysis has principled methods for updating probability distributions. These methods are based on three different, but potentially equally important, sources of information. We demonstrate these three stages using the following concrete problem:

Twelve participants complete a problem-solving task, with only their success or failure measured. In the order that participants completed the problems, the data (D) take the form

$$D = (F,F,F,S,F,F,F,F,S,F,F,S),$$

where F denotes failure and S denotes success.

Undertaking the Analysis

We now present a nontechnical account of the Bayesian analysis of this problem. Technical details are in the appendix.

Information inherent in the problem. Just as zero is the natural starting point for counting, the natural initial representation for Bayesian inference is one corresponding to complete ignorance. Recently, principled methods have been developed for defining prior distributions corresponding to complete ignorance (see Jaynes, 2003, chap. 12). These methods rely on establishing transformational invariances inherent in problems that constrain the choice of prior distribution. Intuitively, the idea is to consider ways in which a problem could be restated, so that it remains fundamentally the same problem but is expressed in a different formal way. Prior distributions must necessarily be invariant under these transformations, because otherwise different ways of stating the same problem would lead to different inferences being drawn. In general, the requirement of invariance from information inherent in a problem provides strong constraints on the choice of prior distribution and often determines the prior distribution uniquely.

As a concrete example, consider using a Weibull distribution to make inferences about response time data. The conclusions drawn by such an analysis obviously should not be affected by the scale of measurement (i.e., expressing the time data in milliseconds rather than seconds should not matter), and so the parameters for the Weibull must have priors that are invariant to changes in scale. As it turns out, this lone constraint is sufficient to determine uniquely the required prior distributions.

For our problem-solving example, complete ignorance about the rate of success is expressed by the prior distribution known as Haldane's prior, which is shown in Figure 1A. A rigorous derivation of this prior using transformational invariance is given by Jaynes (2003, pp. 382–385); an alternative information geometric derivation is provided by Takeuchi (1997); a more intuitive justification is provided by Zhu and Lu (2004). The form of this prior arises because, consistent with the assumption of complete ignorance, it is not known whether the problem-solving task is so

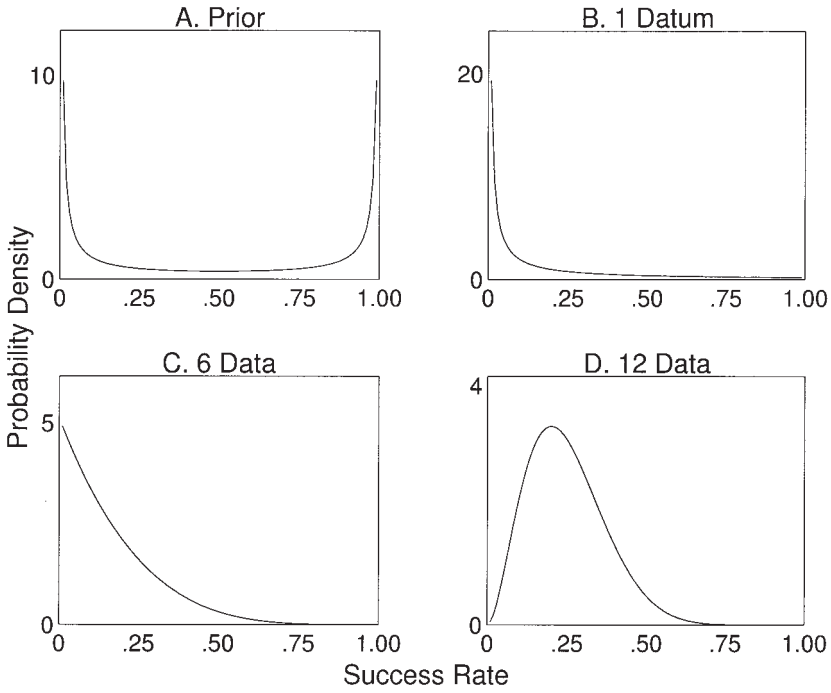

Figure 1. Probability distributions for the rate of success, starting with the completely ignorant prior distribution (Panel A), progressing as 1 (Panel B), 6 (Panel C), and all 12 (Panel D) data are observed.

difficult that it cannot be solved, or so easy that it is always solved. This leads to the extreme possibilities, with success rates of zero or one, being more probable, while still allowing for the possibility that the true rate is somewhere between zero and one. We describe Figures 1B–1D below.

Relevant prior information. If other relevant prior information is available, Bayesian analysis incorporates it into the prior distributions. This is done by updating the prior distributions so that they capture the known constraints provided by additional information but otherwise remain maximally uncertain. For our problem-solving example, a different scenario is that we do know both success and failure are possible. This additional piece of information leads to the uniform prior distribution, which is shown in Figure 2A. Once again, see Jaynes (2003, p. 385) for a derivation.

Information provided by data. As relevant data from experimental or other observations become available, Bayesian analysis uses them to update the probability distribution according to Bayes's theorem. This updating can be done for a complete data set or sequentially as data arrive. Figure 1B–1D and Figure 2B–2D show the updating for 1, 6, and then all 12 data for the problem-solving example, starting with the complete ignorance and uniform priors, respectively. Recall that the first datum is a failure, the first 6 data contain 1 success and 5 failures, and all 12 data have 9 failures and 3 successes.

Comparing the two situations shows the impact the priors, representing different initial states of knowledge, have on our understanding. Under complete ignorance, where it is not known that both successes and failures are possible, the first datum, being a failure, reinforces the possibility that the rate of success might be zero, but eliminates the possibility that the rate is one. In the other scenario, where it is known both are possible, the observation of a failure makes very low rates of success more likely, but still allows some significant likelihood to a range of other possible rates. After 6 data have been observed and a success has been seen, the different states of initial knowledge still influence the distributions, but the influence is less marked. By the time all 12 data have been observed, with a mix of 9 failures and 3 successes, the distributions are very similar.

Statistical Inferences

Standard statistical practice provides essentially two ways of making inferences. One involves parameter estimation, where some quantitative statement is made about the values of unknown parameters of interest. The other involves hypothesis testing, where the probability of a chance-based null account of the data is assessed, to be retained or rejected in favor of an alternative account.

Estimation. The Bayesian equivalent of estimation is to draw conclusions from the posterior distributions of the variables of interest. These conclusions can take any useful form, because the distribution completely describes what is known about the variable. For example, when all 12 data have been observed using the uniform prior, the distribution in Figure 2D allows, among a plethora of others, the following inferences to be made:

- The most likely rate of success (i.e., the mode of the distribution) is .25. This is the single value that is more likely than any other, and so constitutes a best estimate if the interest is in choosing exactly one rate.
- The expected rate of success (i.e., the mean of the distribution) is about .29. This is the value that has the smallest expected

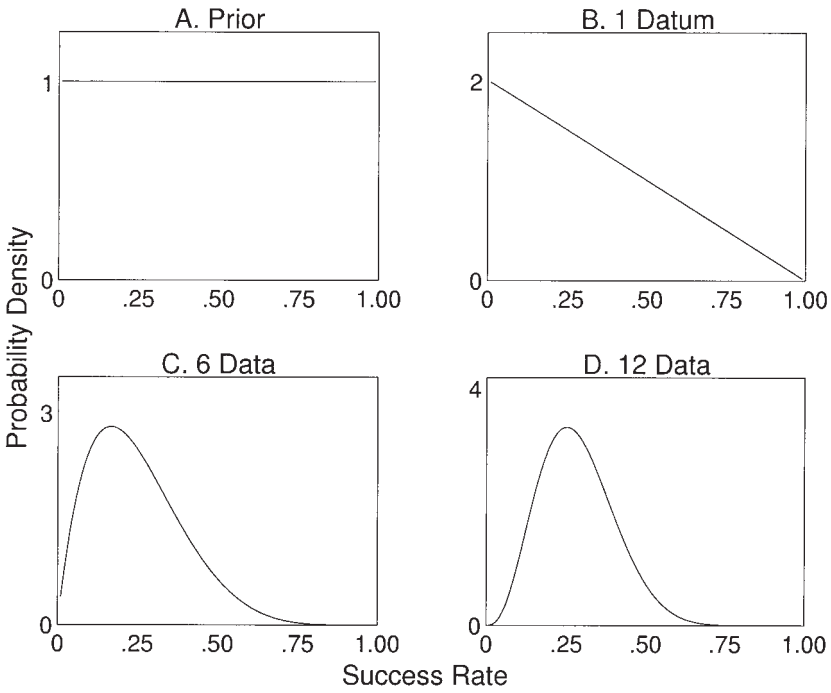

Figure 2. Probability distributions for the rate of success, starting with the uniform prior distribution (Panel A), progressing as 1 (Panel B), 6 (Panel C), and all 12 (Panel D) data are observed.

squared difference from the success rate, and so (in a sense) constitutes a best estimate if the interest is in being as close as possible to the correct value.

• The success rate being .20 is about 2.7 times more likely than the success rate being .10. This is the relative height of the distribution at .20 compared with .10.

• The probability that the success rate is less than .30 is about 59%. This is the area under the distribution that is over the relevant range from 0 to .30.

• The probability that the success rate is between .10 and .55 is about 95%. This is the area under the distribution that is over the relevant range from .10 to .55.

Of course, different answers will be obtained under the different prior assumptions of complete ignorance in Figure 1. This should be expected and is desirable. Indeed, it would be worrisome if a method for statistical inference, when given different information about a problem, did not reach different conclusions. Usually, however, the effects of different prior information are overcome by sufficient data, and so the Bayesian answers to the sorts of questions posed in this list will converge.

Hypothesis testing. The Bayesian equivalent of hypothesis testing is model comparison, where competing accounts (possibly, but not necessarily, including a null account) are compared using Bayes factors (e.g., Kass & Raftery, 1995; Myung & Pitt, 1997). Bayes factors can be thought of as an extension of likelihood ratios, measuring how much more likely one model is than another on the basis of the evidence provided by the data. Applying Bayes factors to our example requires comparing a null hypothesis that assumes a chance success rate with an alternative hypothesis that assumes a success rate different from chance. We assume, for concreteness, that the nature of the problem-solving task means that chance corresponds to a rate of .50, but our method of analysis applies for any other rate.

The Bayes factor is a ratio of two quantities: the probability of the observed data under a certain null hypothesis and the probability of the observed data under the alternative hypothesis. Hence, the Bayes factor quantifies the extent to which the data support the null over the alternative. In our example, it is straightforward to calculate the probability of the observed data (i.e., nine failures and three successes) under a certain null hypothesis specifying a chance success rate of .50. In contrast, it may not be obvious at first sight how to calculate the probability of the data under an alternative hypothesis that assumes the success rate differs from chance, because this composite hypothesis is not associated with any one success rate in particular. Under the Bayesian approach, this probability is calculated as the weighted average of the probability of the data for each individual success rate that differs from chance. The weight given to each possibility in averaging corresponds to its a priori plausibility, as quantified by the prior distribution for the success rate.

As detailed in the appendix, for the uniform prior case shown in Figure 2, the Bayes factor is 1.0 for the prior distribution, and remains at 1.0 after 1 datum has been observed. This means that, at these stages of analysis, the available information does not favor one hypothesis over the other. Once 6 data have been observed, the Bayes factor shows that the alternative hypothesis is about 1.14 times more likely than the null. Once all 12 data are taken into account, the alternative hypothesis has become 1.43 times more likely.

Of course, experimental psychologists are often interested in comparing two groups, rather than comparing one group against a null hypothesis. Suppose, for example, a second smaller group of

advanced problem solvers is observed to solve 5 of 6 problems, and an inference is required about whether 5 out of 6 corresponds to the same success rate as the original sample performance of 9 out of 12. Bayesian inference also handles this problem by using Bayes factors. For this example, it is about 6.9 times more likely that the advanced group will have a greater success rate than the original control group. As with estimation, both this Bayes factor and the one comparing the original group to the null hypothesis, will differ under the prior assumption of complete ignorance, but will generally converge as the evidence provided by data dominates prior information.

Discussion

The two quantities highlighted by Trafimow (2003) as being problematic for Bayesian analysis—the prior probability of the null hypothesis, and the probability of the data when the null hypothesis is false—are, in fact, both readily calculated using priors corresponding to complete ignorance.[1] To be fair, the definition of *complete-ignorance priors* in terms of transformational invariance is a recent development in Bayesian statistics and addresses what had previously been a serious deficiency. Previously, the definition of priors had relied on applying the so-called *principle of indifference* or *principle of insufficient reason*, often associated with Laplace (e.g., Laplace, 1951), to the events whose relative probabilities were being quantified by the prior distribution. This principle simply assumes, in the absence of other evidence, that all possibilities are equally likely. As has repeatedly been pointed out (e.g., A. Edwards, 1972; Fisher, 1930, 1932, 1956), including by those advocating the Bayesian approach (e.g., Kass & Wasserman, 1996), this principle is highly problematic when applied to events, because different (but equally reasonable) ways of partitioning a variable into a set of events lead to different inferences being drawn.[2] One way of understanding transformational invariance is that it applies the principle of indifference at the level of problems rather than events, where there is no such arbitrariness, because it relies only on the explicit statement of the problem.[3]

With the advent of an objective method for determining priors, our rate example shows the feasibility of Bayesian inference for psychology in a concrete way. Whether one possible rate is more likely than another can be assessed, and how much more likely it is can be quantified. The probability that the rate is greater than some value or is between two values can be measured. Observed data can be compared with chance or with data from another experimental group. Whatever the question, the important point is that Bayesian inferences are always possible in principle. Even in the most extreme case, where the problem has been defined, but absolutely nothing else is known, Figure 1A shows the distribution on which inferences should be based.

Our example also provides several demonstrations that Bayesian analysis is sometimes necessary. One of these situations is where there is relevant prior information. Figure 2A is different from Figure 1A because additional relevant prior information is available. Standard methods, with their reliance on sampling distributions of data, have no easy way of incorporating prior information. Another obvious advantage of Bayesian analysis in our example is its ability to handle small numbers of data. Many routinely applied estimators are large-sample approximations. An attraction of the Bayesian approach is that the same method can be applied validly to any sample size. The same holds true for hypothesis testing, where Bayesian methods always work but ones based on NHST or confidence intervals can be inefficient or pathological. Indeed, it is well documented (e.g., Carlin & Louis, 2000; Lindley & Phillips, 1976) that NHST and confidence interval methods behave incoherently for our example, because they violate what is known as the likelihood principle. Our goal is to explain Bayesian methods rather than criticize standard ones, so we encourage the interested reader to consult Berger and Wolpert (1984) and W. Edwards, Lindman, and Savage (1963) for details.

Of course, our example is one where Bayesian methods are computationally as well as conceptually straightforward. However, modern computing capabilities, and the numerical algorithms that use them (e.g., Gelman, Carlin, Stern, & Rubin, 1995; Gilks, Richardson, & Spiegelhalter, 1996), have rendered less elegant inference problems equally tractable. These algorithms are progressively being incorporated into standard statistical software packages.[4] Bayesian inference maintains its conceptual simplicity throughout, but more computational work is required to generate posterior distributions and Bayes factors.

We would argue that, despite the temptation of retaining current methods, making the additional computational effort to use Bayesian methods will often be rewarded. Psychology has many important and interesting problems with characteristics that do not suit current methods, but can be dealt with in a principled and unified way by the Bayesian approach. This is why Bayesian inference deserves serious study and widespread adoption in psychology.

[1] The averaging procedure used to calculate the probability of the data under a composite hypothesis follows from elementary probability theory. Therefore, the claim that it is impossible to calculate the probability of the observed data when the null hypothesis is false reduces to the argument that it is not possible to determine prior distributions.

[2] We note that unbiased estimators, efficient estimators, shortest confidence intervals, and other key quantities in standard statistics continue to suffer from the lack of reparameterization invariance that is at the heart of this problem (Jaynes, 2003, p. 377). If objections to the principle of indifference are Trafimow's (2003) justification for rejecting Bayesian inference, then standard statistics certainly must also be rejected.

[3] A concrete example of this key difference is Jaynes's (2003, pp. 386–394) solution of the Bertrand paradox—a problem that had previously provided one of the best examples of the difficulties with the principle of indifference—using transformational invariance to determine the required prior distributions.

[4] For example, Systat (Version 11; 2002) has general Markov chain Monte Carlo sampling procedures.

References

Berger, J. O., & Wolpert, R. L. (1984). *The likelihood principle*. Hayward, CA: Institute of Mathematical Statistics.

Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis* (2nd ed.). New York: Chapman & Hall.

Cox, R. T. (1961). *The algebra of probable inference*. Baltimore: Johns Hopkins University Press.

Edwards, A. W. F. (1972). *Likelihood*. Cambridge, England: Cambridge University Press.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society, 26,* 528–535.

Fisher, R. A. (1932). Inverse probability and the use of likelihood. *Proceedings of the Cambridge Philosophical Society, 28,* 257–261.
Fisher, R. A. (1956). *Statistical methods and scientific inference.* Edinburgh, Scotland: Oliver & Boyd.
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London: Chapman & Hall.
Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice.* London: Chapman & Hall.
Gregory, P. C. (1999). Bayesian period signal detection: I. Analysis of 20 years of radio flux measurement of the x-ray binary LSI+61 303. *Astrophysical Journal, 520,* 361–375.
Gull, S. F., & Daniell, G. J. (1980). The maximum entropy algorithm applied to image enhancement. *Proceedings of the IEEE, 5,* 170.
Jaynes, E. T. (2003). *Probability theory: The logic of science.* New York: Cambridge University Press.
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.
Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91*(435), 1343–1370.
Laplace, P.-S. (1951). *A philosophical essay on probabilities* (F. W. Truscott & F. L. Emory, Trans.). New York: Dover. (Original work published 1814)
Lindley, D. V. (1972). *Bayesian statistics: A review.* Philadelphia: Society for Industrial and Applied Mathematics.
Lindley, D. V., & Phillips, L. D. (1976). Inference for a Bernoulli process (a Bayesian view). *American Statistician, 30,* 112–119.
Myung, I. J., Forster, M., & Browne, M. W. (2000). A special issue on model selection. *Journal of Mathematical Psychology, 44,* 1–2.
Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review 4*(1), 79–95.
Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472–491.
Press, S. J. (1989). *Bayesian statistics: Principles, models, and applications.* New York: Wiley.
Sivia, D. S., Carlile, C. J., & Howells, W. S. (1992). Bayesian analysis of quasielastic neutron scattering data. *Physica B, 182,* 341–348.
Systat (Version 11) [Computer software]. (2002). Richmond, CA: Systat Software.
Takeuchi, J. (1997). Characterization of the Bayes estimator and the MDL estimator for exponential families. *IEEE Transactions on Information Theory, 43,* 1165–1174.
Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review, 110,* 526–535.
Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review 7*(3), 424–465.
Zellner, A. (1984). *Basic issues in econometrics.* Chicago: University of Chicago Press.
Zhu, M., & Lu, A. Y. (2004). The counter-intuitive non-informative prior for the Bernoulli family. *Journal of Statistics Education* [Electronic version], *12*(2). Retrieved October 6, 2004, from http://www.amstat.org/publications/jse/v12n2/zhu.pdf

Appendix

Technical Details

This appendix provides a formal treatment of the worked example of Bayesian analysis. The complete ignorance prior is found by Jaynes (2003, pp. 382–385) as the probability distribution that is invariant under the transformation $s' = as/(1 - s + as)$, where the rate of success $s \in [0, 1]$ and $a > 0$. The unique solution is known as Haldane's prior, and is given by $p_H(s) \propto 1/[s(1 - s)]$. Jaynes then showed that updating this prior to include the knowledge that both successes and failures are possible leads to the uniform prior, given by $p_U(s) = 1$.

Both of these prior states of knowledge can conveniently be represented as beta distributions

$$\text{Beta}_{(a,\,b)}(s) = \frac{s^{a-1}(1-s)^{b-1}}{B(a,\,b)},$$

where $B(a, b) = (a-1)!(b-1)!/(a+b-1)!$ is the beta function, with $a = b = 0$ for p_H and $a = b = 1$ for p_U.

For any available data D showing k successes out of n trials, Bayes's theorem gives the posterior distribution

$$p(s|D) = \text{Beta}_{(b+k,\,b+n-k)}(s).$$

Every distribution shown in Figures 1 and 2 corresponds to the appropriate beta distribution (i.e., the relevant prior, updated by the available data).

Hypothesis testing with the null and alternative models, H_0 and H_1, is based on the Bayes factor. For our example, the Bayes factor is

$$\frac{p(D|H_0)}{p(D|H_1)} = \frac{\binom{n}{k}\left(\frac{1}{2}\right)^n}{\int_0^1 \binom{n}{k} s^k(1-s)^{n-k}\,\text{Beta}_{(b,\,b)}(s)\,ds}$$

$$= \frac{\left(\frac{1}{2}\right)^n}{B(b+k,\,b+n-k)}. \quad (1)$$

Note that the denominator in Equation 1 is the probability of the data given that the null is false (i.e., given that the interval alternative hypothesis is true). This is one of the quantities that Trafimow (2003) incorrectly argued is not available for Bayesian analysis.

Using Equation 1, the Bayes factors reported for our example were found simply by substituting the appropriate numbers of observations and successes, using the appropriate prior. For example, Figure 2A corresponds to $k = 0$ and $n = 0$ with the prior having $b = 1$. Substituting these numbers into Equation 1 gives the reported Bayes factor of 1.0. As a second example, Figure 2D corresponds to $k = 3$, $n = 12$, and $b = 1$. Substituting these numbers gives a Bayes factor of approximately .70 in favor of the null hypothesis. This corresponds to a Bayes factor of $1/.70 \approx 1.43$ in favor of the alternative hypothesis, as reported in the text.

The reported comparison between the control and advanced groups was done by applying the Bayes factor that compares a same-rate model M_s against a different-rates model M_d. For two groups respectively having k_1 and k_2 successes out of n_1 and n_2 trials, the same-rate model assumes the same success rate (s) applies for both groups. The different-rates model assumes different success rates (s_1 and s_2) for the two groups. The Bayes factor comparing these models is given by

$$\frac{p(D|M_s)}{p(D|M_d)} = \frac{\int_0^1 p(D|s)p(s)ds}{\int_0^1 \int_0^1 p(D|s_1, s_2)\, p(s_1, s_2)ds_1ds_2}$$

$$= \frac{\int_0^1 \binom{n_1}{k_1} s^{k_1}(1-s)^{n_1-k_1} \binom{n_2}{k_2} s^{k_2}(1-s)^{n_2-k_2}\, \mathrm{Beta}_{(b,b)}(s)ds}{\int_0^1 \int_0^1 \binom{n_1}{k_1} s_1^{k_1}(1-s_1)^{n_1-k_1}\mathrm{Beta}_{(b,b)}(s_1) \binom{n_2}{k_2} s_2^{k_2}(1-s_2)^{n_2-k_2}\mathrm{Beta}_{(b,b)}(s_2)ds_1ds_2}$$

$$= \frac{B(b+k_1+k_2,\, b+n_1+n_2-k_1-k_2)}{B(b+k_1,\, b+n_1-k_1)B(b+k_2,\, b+n_2-k_2)}. \quad (2)$$

For our example, the counts of the numbers of successes and total trials for each group are $k_1 = 3$, $n_1 = 12$, $k_2 = 5$, and $n_2 = 6$. Substituting these values, together with $b = 1$ for the uniform prior, into Equation 2 gives a Bayes factor of approximately .145 in favor of the same-rate model. This corresponds to the Bayes factor of $1/.145 \approx 6.9$ in favor of the different-rates model reported in the text.

Received June 28, 2004
Revision received October 6, 2004
Accepted October 7, 2004 ■

Postscript: Bayesian Statistical Inference in Psychology: Comment on Trafimow (2003)

Michael D. Lee
University of Adelaide

Eric-Jan Wagenmakers
University of Amsterdam

Trafimow (2005) is right in pointing out that our comment (Lee & Wagenmakers, 2005) focused on using Bayesian methods to make statistical inferences about psychological models and data, at the neglect of philosophical issues. And there are certainly some philosophical issues that are worth discussing. It seems, for example, that our comment should have made it clear that the objective Bayesian approach we advocate views probabilities neither as relative frequencies nor as belief states, but as degrees of plausibility assigned to propositions in a rational way, entirely (and uniquely) determined by the available information. Harder philosophical problems that we think are important, but did not mention, are whether it is possible to give a complete statistical characterization for any data-generating process and how incompleteness results (such as Gödel's theorem) might apply to the Bayesian framework.

Unfortunately, these substantial issues are not addressed. Instead, Trafimow (2005) provides a series of consternations making it clear that Bayesian methods cannot work miracles for psychological researchers and that researchers must be resigned to using them to provide coherent and rational analyses of available information. The Fisher example shows that Bayesian inference cannot divine true information with which it was never provided. The amazing theory of memory example and the dating example show that Bayesian inference cannot resolve the indecision of researchers about what hypothesis they are testing and what information they have available. And, while we are not sure we understand the point of the Suppes example, we agree that Bayesian inference cannot account for the quirks of human language production under uncertainty. In the end, these consternations seem to leave Trafimow troubled by all methods for analyzing psychological models and data. Perhaps he is right, and Bayesians should "down tools," waiting for a method of inference that is not only coherent and rational, but can also work miracles. But, even if we had fortunes, we wouldn't bet them on it.

References

Lee, M. D., & Wagenmakers, E.-J. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review, 112,* 662–668.

Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review, 110,* 526–535.

Trafimow, D. (2005). The ubiquitous Laplacian assumption: Reply to Lee and Wagenmakers (2005). *Psychological Review, 112,* 669–674.

Received June 28, 2004
Revision received October 6, 2004
Accepted October 7, 2004 ■