That Wretched Prior: Online Appendix to "A Practical Solution to the Pervasive Problems of p–Values", to appear in *Psychonomic Bulletin & Review*

Eric-Jan Wagenmakers
University of Amsterdam

Correspondence concerning this online appendix should be addressed to:
Eric–Jan Wagenmakers
University of Amsterdam, Department of Psychology
Roetersstraat 15
1018 WB Amsterdam, The Netherlands
Ph: (+31) 20–525–6876
Fax: (+31) 20-639-0279
E-mail may be sent to EJ.Wagenmakers@gmail.com.

This appendix discusses the issue of priors in some more detail (cf. Lindley, 2004). Basically, priors can be determined by two different methods. The first method is known as "subjective", and it is the one that has the fewest computational and philosophical problems. A "subjective Bayesian" argues that all inference is necessarily relative to a particular state of knowledge. This state of knowledge differs from one individual to the next. The claim is that objective knowledge of the physical world is an unattainable, quixotic ideal. In this sense then, "probability does not exist" (de Finetti, 1974). For a subjective Bayesian, the prior simply quantifies her personal degree of belief that is to be adjusted by the data.

In an article entitled "Why isn't everyone a Bayesian?", Efron argued that one of the major drawbacks of the subjective Bayesian paradigm is the fact that it does not lend itself well for scientific communication: "Strict objectivity is one of the crucial factors separating scientific thinking from wishful thinking." (Efron, 1986, p. 3). I suspect that many experimental psychologists would agree with Efron. It may appear that by opening a Pandora's box of subjective priors, one can pick and choose the prior that allows one to bias the process of scientific inference at will.

Several arguments can be brought to bear against Efron's claim, however. First, it is not clear what it means to be "objective". As pointed out by Berger (1985, p. 125), "(...) when different reasonable priors yield substantially different answers, can it be right to state that there *is* a single answer? Would it not be better to admit that there is scientific uncertainty, with the conclusion depending on prior beliefs?". Second, the fact that a procedure is "objective" provides no guarantee that the procedure is rational or coherent. Third, this article contains many well–known examples that demonstrate the frequentist procedures are also subjective (cf. Berger & Berry, 1988). Frequentist procedures depend on hypothetical actions for imaginary events. This hardly provides much support for the statement that “The high ground of scientific objectivity has been seized by the frequentists.” (Efron, 1986, p. 4). In fact, a Bayesian might argue that the subjectivity in the Bayesian analysis is in the prior, which is completely specified and available for scientific scrutiny. In contrast, the subjectivity in the frequentist paradigm is hidden, as it pertains to what the researcher was thinking as she collected the data. It would therefore be more accurate to state that “In some fields, researchers mistakenly believe that the high ground of scientific objectivity has been seized by the frequentists.”.

The second method to specify priors is by means of an “objective” Bayesian analysis (Kass & Wasserman, 1996). What this means is that priors are specified according to certain predetermined rules. For instance, one may use the unit information prior (i.e., a prior that carries as much information as a single observation, Kass & Wasserman, 1995), a prior that is invariant under transformations (Jeffreys, 1961) or a prior that maximizes entropy (Jaynes, 1968). Objective priors are generally vague or uninformative, that is, thinly spread out over the range for which they are defined. Often, an objective prior is chosen so as to reflect a state of ignorance. This is very different from the subjective Bayesian approach, in which priors are chosen so as to reflect personal degree of belief.

The objective Bayesian approach has several advantages over the subjective approach. First of all, most researchers will find it appealing that statistical inference is independent of the person who performs the analysis. Of course, as with frequentist procedures, this objectivity is only partial, as different researchers might use different models to analyze the same data. For instance, researcher A might use hierarchical linear regression, researcher B might use a regular analysis of variance after transforming the data to achieve normality, researcher C might first remove outliers from the data, etc. As argued by Box “(...) it is impossible logically to distinguish between model assumptions and the prior distribution of the parameters.” (Box, 1980, p. 384). Thus, I take objectivity to mean that given the same data *and* the same assumptions regarding the model, different researchers will arrive at the same conclusions. The second advantage of the objective approach is that in models with very many parameters, it foregoes the need to consider one’s personal beliefs for every single parameter.

From a pragmatic perspective, the discussion of priors would be moot if it could be shown – either in general or in specific problems – that the specific shape of the prior did not greatly affect inference (cf. Dickey, 1973). Consider Bayesian inference for the mean μ of a normal distribution. For parameter estimation, one can specify a prior $Pr(\mu)$ that is very uninformative (e.g., spread out across the entire real line). The data will quickly overwhelm the prior, and hence parameter estimation is relatively robust to the specific choice of prior. In contrast, the Bayes factor for a two–sided hypothesis test is sensitive to the shape of the prior (Lindley, 1957; Shafer, 1982). This is not surprising – if we increase the interval along which μ is allowed to vary according to H_1, we effectively increase the complexity of H_1. The inclusion of unlikely values for μ decreases the average likelihood for the observed data. For a subjective Bayesian, this is not really an issue, as $Pr(\mu)$ reflects her prior belief. For an objective Bayesian, hypothesis testing constitutes a bigger challenge: On the one hand, the prior needs to be vague, as it reflects a state of ignorance. On the other hand, a prior that is too vague can increase the complexity of H_1 to such an extent that H_1 will always have low posterior probability, regardless of the observed data. Several objective Bayesian procedures have been developed that try to address this dilemma, examples including the local Bayes factor (Smith & Spiegelhalter, 1980), the intrinsic Bayes factor (Berger & Pericchi, 1996), the partial Bayes factor (O'Hagan, 1997) and the fractional Bayes factor (O'Hagan, 1997) (for a summary see Gill, 2002, Chapter 7).

It is important to realize that to many Bayesians the presence of the prior is actually an asset rather than a nuisance. First of all, the prior ensures that different sources of information are appropriately combined, such as when the posterior after observation of a batch of data D_1 becomes the prior for the observation of a new batch of data D_2. If the data are conditionally independent, the Bayesian analysis will arrive at the same conclusion regardless of the temporal order or data (i.e., D_1 first and D_2 second or vice versa) and regardless of whether the data arrived one–by–one, in batches, or all at once. Mathematical derivation shows that in order to avoid making irrational and inconsistent decisions, the quantification of uncertainty needs to obey the rules of probability theory – this includes Bayes rule, which in turn includes the priors (for details see Bernardo & Smith, 1994; Cox, 1946; D'Agostini, 1999; Fishburn, 1986; Jaynes, 2003; Jeffreys, 1961; Lindley, 1982, 2004; for a discussion see Colyvan, 2004; Van Horn, 2003). In Bayesian statistics, irrational behavior goes under the name of *incoherence*. In the context of betting, de Finetti showed that incoherence makes someone a sure loser (de Finetti, 1974; see also Smith, 1961, and Cornfield, 1969).[1]

A second advantage of the prior is that it can prevent one from making extreme and implausible inferences; the prior may "shrink" the extreme estimates toward more plausible values. To illustrate, Rouder, Lu, Speckman, Sun, and Jiang (2005) discuss a baseball game in which the struggling Kansas City Royals lead the favorite Boston Red Sox by 5–0 after one inning. Based on a maximum likelihood approach, the expected final score after this one inning is 5×9 innings $= 45 - 0$ in favor of the Kansas City Royals. To anyone with even the slightest knowledge of the game, this score is wildly implausible. Other examples of why priors can be helpful for inference are given by Box and Tiao (1973, pp. 19–20) and Lindley and Phillips (1976).

A third advantage of specifying priors is that it allows one to focus on parameters of interest by eliminating so–called nuisance parameters through the law of total probability. Let θ be a parameter of interest, and let γ be a nuisance parameter. To obtain the posterior distribution of θ, we simply integrate the joint posterior $Pr(\theta, \gamma|D)$ over all values of γ, that is, $Pr(\theta|D) = \int_\Gamma Pr(\theta, \gamma|D)d\gamma = \int_\Gamma Pr(\theta|\gamma, D)Pr(\gamma)d\gamma$. This method is very general and can often be tremendously useful.

These considerations suggest that one might turn the problem on its head: What does one make of inferential procedures that are incapable of taking prior knowledge into account? Such procedures will be surely be incoherent (Lindley, 1977), may waste useful information, and may lead to implausible estimates. In this context, Lindley notes that "The statistician who reports a confidence interval of $(-0.5, 2.3)$ for a parameter is ridiculous in the opinion of an investigator who knows the parameter is positive." (Lindley, 2004, p. 85). A Bayesian could easily incorporate this knowledge through the prior, by assigning probability mass to positive values only. Jaynes (2003, p. 373) states the case for the prior even more strongly: "If one fails to specify the prior information, a problem of inference is just as ill–posed as

[1]Unfortunately, a discussion of de Finetti's definition of probability as fair betting odds would take us too far afield. The interested reader is referred to numerous internet resources for more information (e.g., `http://en.wikipedia.org/wiki/Bruno_de_Finetti`).

if one had failed to specify the data.”.

References

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer.

Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, *76*, 159–165.

Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*, 109–122.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* New York: Wiley.

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society, Series A*, *143*, 383-430.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis.* Reading: Addison–Wesley.

Colyvan, M. (2004). The philosophical significance of Cox's theorem. *International Journal of Approximate Reasoning*, *37*, 71-85.

Cornfield, J. (1969). The Bayesian outlook and its application. *Biometrics*, *25*, 617–657.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *The American Journal of Physics*, *14*, 1–13.

D'Agostini, G. (1999). Teaching statistics in the physics curriculum: Unifying and clarifying role of subjective probability. *American Journal of Physics*, *67*, 1260–1268.

de Finetti, B. (1974). *Theory of probability, vol. 1 and 2.* New York: John Wiley & Sons.

Dickey, J. M. (1973). Scientific reporting and personal probabilities: Student's hypothesis. *Journal of the Royal Statistical Society B*, *35*, 285–305.

Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician*, *40*, 1–5.

Fishburn, P. C. (1986). The axioms of subjective probability. *Statistical Science*, *1*, 335–345.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach.* Boca Raton (FL): CRC Press.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, *4*, 227–241.

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge, UK: Cambridge University Press.

Jeffreys, H. (1961). *Theory of probability.* Oxford, UK: Oxford University Press.

Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, *90*, 928–934.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, *91*, 1343–1370.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, *44*, 187–192.

Lindley, D. V. (1977). The distinction between inference and decision. *Synthese*, *36*, 51–58.

Lindley, D. V. (1982). Scoring rules and the inevitability of probability. *International Statistical Review*, *50*, 1–26.

Lindley, D. V. (2004). That wretched prior. *Significance*, *1*, 85–87.

Lindley, D. V., & Phillips, L. D. (1976). Inference for a Bernoulli process (a Bayesian view). *The American Statistician*, *30*, 112–119.

O'Hagan, A. (1997). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society B*, *57*, 99–138.

Rouder, J. N., Lu, J., Speckman, P. L., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, *12*, 195-223.

Shafer, G. (1982). Lindley's paradox. *Journal of the American Statistical Association*, *77*, 325–351.

Smith, A. F. M., & Spiegelhalter, D. J. (1980). Bayes factors and choice criteria for linear models. *Journal of the Royal Statistical Society B*, *42*, 213–220.

Smith, C. A. B. (1961). Consistency in statistical inference and decision (with discussion). *Journal of the Royal Statistical Society B*, *23*, 1–37.

Van Horn, K. S. (2003). Constructing a logic of plausible inference: A guide to Cox's theorem. *International Journal of Approximate Reasoning*, *34*, 3-24.