



*Author response*

## Rejoinder to discussion of ‘Philosophy and the practice of Bayesian statistics’

Andrew Gelman<sup>1\*</sup> and Cosma Shalizi<sup>2</sup>

<sup>1</sup>Department of Statistics and Department of Political Science, Columbia University, New York, USA

<sup>2</sup>Department of Statistics, Carnegie Mellon University, Pittsburgh, USA

### Different views of Bayesian inference

The main point of our paper was to dispute the commonly held view that Bayesian statistics is or should be an algorithmic, inductive process culminating in the calculation of the posterior probabilities of competing models. Instead, we argued that effective data analysis – Bayesian or otherwise – proceeds more messily through a jagged process of formulating research hypotheses, exploring their implications in light of data, and rejecting aspects of our models in light of systematic misfits compared to available data or other sources of information. We associate this last bit with Popper’s falsification or (weakly) with Kuhn’s scientific revolutions, but these connections with classical philosophy of science are not crucial. Our real point is that Bayesian data analysis, in the form that we understand and practise, requires the active involvement of the researcher in constructing and criticizing models, and that from this perspective the entire process of Bayesian prior-to-posterior inference can be seen as an elaborate way of understanding the implications of a model so that it can be effectively tested. Just as a chicken is said to be nothing but an egg’s way of making another egg, so posterior inference is a way that a model can evaluate itself. But this inference and evaluation process, in our view, has essentially nothing to do with calculations of the posterior probability of competing models. As we discuss in our paper, for technical, philosophical, and historical reasons we tend not to trust such marginal posterior probabilities. (See Figure 1 of our paper for the sort of reasoning that we do *not* like.)

Given all this, the discussion of our paper is remarkably uncontentious: none of the five discussants express support for the standard (according to Wikipedia) view of an overarching inductive Bayesian inference, and all agree with us that the messiness of

---

\*Correspondence should be addressed to Andrew Gelman, Department of Statistics and Department of Political Science, Columbia University, New York, NY 10027, USA (e-mail: gelman@stat.columbia.edu).

real-world data analysis is central to statistical reasoning, not a mere obstacle to be cleaned up by means of a better prior distribution.

The discussants present different perspectives, but a common theme is that our own recommended approach of Bayesian analysis and posterior predictive checking is itself limited or, at the very least, is only one of many ways to approach statistical inference and decision-making.

We agree, and we briefly respond to each discussant in turn and then summarize our points.

## Response to specific comments

Denny Borsboom (2013) points out that there are other Bayesian philosophies beyond the two discussed in our paper. We considered the ‘usual story’ based on computing the posterior probabilities of competing models and our preferred falsificationist attitude. Borsboom argues that a fuller philosophy of statistics – Bayesian or otherwise – should also account for confirmation and construction of models as well as inference and criticism of models that have already been proposed. We agree that our philosophy is incomplete and welcome such additions. We have had ideas of models for the model-building process using a recursive language-like framework in which a model is built from existing pieces (by analogy with the stepwise curve-fitting algorithm of Schmidt & Lipson, 2009; and reviewed by Gelman, 2009). But such models are a distant approximation to how we actually put models together. We hope that the philosophical approaches suggested by Borsboom lead us to a better understanding of the interaction between inference and the construction of models.

John Kruschke (2013) argues in the opposite direction, that our philosophy is not Bayesian enough, and that with a careful re-expression we can integrate predictive model checking into the inductive Bayesian fold. Kruschke notes that the act of interpreting a model check is itself a form of inference, perhaps with our brain’s visual system performing some version of Bayesian decision-making. Indeed, one way to conceptualize the incompleteness of our philosophical framework is to imagine trying to program Bayesian data analysis in an artificial intelligence system. Inference would be no problem (at least for the large class of models that can be fitted in reasonable time using Markov chain Monte Carlo, variational Bayes, or some other existing computational approach). And we could just about imagine model expansion being performed algorithmically using some alphabet of models. But how would the artificial intelligence perform posterior predictive checks, if the program does not have a ‘homunculus’ to assess discrepancies between observed and predicted data? Kruschke is perhaps correct that this sort of comparison could itself be performed Bayesianly, and we are interested in the potential of this process being automated.<sup>1</sup> Deborah Mayo (2013) explains how, in our efforts to model our modelling process, we have oversimplified the philosophies of Popper and others. Here we fear we fall into a long tradition of scientists who attempt to develop philosophical principles via introspection – but without introspecting carefully enough.

---

<sup>1</sup> We disagree, however, with Kruschke’s view that it is desirable to penalize complex models, automatically or otherwise. Instead we prefer the following dictum from Radford Neal (1996): Sometimes a simple model will outperform a more complex model . . . Nevertheless, I [Neal] believe that deliberately limiting the complexity of the model is not fruitful when the problem is evidently complex. Instead, if a simple model is found that outperforms some particular complex model, the appropriate response is to define a different complex model that captures whatever aspect of the problem led to the simple model performing well.

This is a good place for us to repeat our belief and hope that we are clearing the air by describing how we do Bayesian inference without comparing the posterior probabilities of models. Describing the philosophy of what we *do* do – that is more of a challenge. In particular, Mayo picks up on a hole in our philosophy that matches a similar gap in classical statistics: how do we decide when a discrepancy between replications and data is ‘statistically significant’, and what do we do about it? It is all well and good for us to emphasize practical significance and our concern for aspects of model misfit that are substantively important, but we still end up looking at  $p$ -values (or their graphical equivalent) one way or another. We are still struggling with this issue in our applied work: this gap in our philosophy represents a gap in our practical tools as well.

Richard Morey and colleagues (2013) argues that we are too quick to abandon Bayesian philosophy: by considering priors as ‘regularization devices’ rather than as true probability distributions, we are abandoning ‘the interpretation of the corresponding posterior’ and thus diminishing the value of the simulations of predictive data that we are using to check our model. In our paper we frame the strong assumptions of Bayesian inference as a feature rather than a bug: the stronger the assumptions, the more ways the model can be checked. This is a Popperian idea, that the best models make lots of predictions and are ready to be refuted, with the act of refutation being the spur to improvement. Morey, Romeijn, and Rouder is going one more step, noting that to the extent that we equivocate about the probabilistic nature of our priors, we are reducing our ability to learn from falsification. Belief is the foundation of scepticism, and by refusing to commit we are also losing an opportunity to refute. In that spirit, Morey, Romeijn, and Rouder would like to preserve the marginal probability calculation giving the relative (although not absolute) posterior probabilities of competing models, thus taking a half-way point between the standard view (in which new evidence causes the better model to dominate, with no need for the steps of model checking and improvement) and our view (in which these marginal probabilities and Bayes factors are so dependent on arbitrary aspects of the model as to be useless; see Section 4.3 of our paper).

Stephen Senn (2013) likes our applied work but points out some holes in our theory. This is important because undoubtedly we could have achieved similar results using other statistical approaches. Bayes is fine but other regularization methods could also do the job. In practice the following seem to be important in developing a method to solve problems in applied statistics: (1) the method must have a way to incorporate diverse sources of data (e.g., survey responses, demographic information, and election outcomes in the vote modelling problem); (2) when large amounts of data come in, the method must be flexible enough to expand, either non-parametrically or through a sieve-like set of increasingly dense forms; and (3) the estimation must be regularized to avoid overfitting. Bayesian inference has some particular advantages in that it automatically unifies inference and prediction (and its predictive simulations can be directly used to check model fit), but these are second-order benefits. Other modes of inference can be hacked to yield probabilistic predictions as needed. In any case, Senn points out a problem in our philosophy as well as other formulations of realistic statistical practice: we choose our model based on its fit to the data, thus the statistical properties of the *model* we choose are not the same as the (generally unstated) statistical properties of our full *procedure*. We do not know how important this is, but we suppose that a useful start would be to investigate this difference in some special cases in which a class of models is fitted and some rule is used to stop or go forward (we will not say ‘accept or reject’) given the results of a posterior predictive check.

As can be seen from our comments, the discussants raise complementary points. In our philosophy, neither model building nor model checking is fully formulated. Our weak defence is that in practice these steps are not so well understood but are part of any serious applied modelling, thus we prefer to include model building and checking as open-ended steps. We want our framework to catch up with statistical practice, but we find it difficult to devise a philosophy that anticipates future methods. But this is only a weak response: all the discussants raise important ideas that point towards potentially useful research in modelling the process of applied statistics.

As we say in our paper, the philosophy of statistics is not a mere game: wrong philosophies can trap people in relatively ineffective methods (this is how we feel about many of the applications of Bayes factors), whereas forward-looking philosophies can point towards methodological improvements (such as the ideas for Bayesian model building and model checking raised by some of the discussants here).

### Looking forward

In summary, our goal in writing our paper was not to say that Bayesian inference is ‘better’ but to delineate what is done when we do Bayes, as compared to the ‘party line’ of what people say is done.

When we were beginning our statistical educations, the word ‘Bayesian’ conveyed membership in an obscure cult. Statisticians who were outside the charmed circle could ignore the Bayesian subfield, while Bayesians themselves tended to be either apologetic or brazenly defiant. These two extremes manifested themselves in ever more elaborate proposals for non-informative priors, on the one hand, and declarations of the purity of subjective probability, on the other.

Much has changed in the past 30 years. ‘Bayesian’ is now often used in casual scientific parlance as a synonym for ‘rational’, the anti-Bayesians have mostly disappeared, and non-Bayesian statisticians feel the need to keep up with developments in Bayesian modelling and computation. Bayesians themselves feel more comfortable than ever constructing models based on prior information without feeling an obligation to be non-parametric or a need for priors to fully represent a subjective state of knowledge.

In short, Bayesian data analysis has become normalized. Our paper is an attempt to construct a philosophical framework that captures applied Bayesian inference as we see it, recognizing that Bayesian methods are highly assumption-driven (compared to other statistical methods) but that such assumptions allow more opportunities for a model to be checked, for its discrepancies with data to be explored.

We felt that a combination of the ideas of Popper, Kuhn, Lakatos, and Mayo covered much of what we were looking for – a philosophy that combined model building with constructive falsification – but we recognize that we are, at best, amateur philosophers. Thus we feel our main contribution is to consider Bayesian data analysis worth philosophizing about.

Bayesian methods have seen huge advances in the past few decades. It is time for Bayesian philosophy to catch up, and we see our paper as the beginning, not the end, of this process.

## Acknowledgements

We thank the editors of this journal for organizing the discussion and the US National Science Foundation for partial support of this work.

## References

- Borsboom, D. (2013). How to practise Bayesian statistics outside the Bayesian church: what philosophy for Bayesian statistical modelling? *British Journal of Mathematical and Statistical Psychology*, *66*, 39–44. doi:10.1111/j.2044-8317.2011.02062.x
- Gelman, A. (2009). Equation search, part 2. Statistical Modeling, Causal Inference, and Social Science blog, 8 December. Retrieved from [http://andrewgelman.com/2009/12/equation\\_search\\_1/](http://andrewgelman.com/2009/12/equation_search_1/)
- Kruschke, J. (2013). Posterior predictive checks can and should be Bayesian: comment on Gelman and Shalizi, 'Philosophy and the practice of Bayesian statistics'. *British Journal of Mathematical and Statistical Psychology*, *66*, 45–56. doi:10.1111/j.2044-8317.2011.02063.x
- Mayo, D. (2013). The error-statistical philosophy and the practice of Bayesian statistics: comments on Gelman and Shalizi: philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*, 57–64. doi:10.1111/j.2044-8317.2011.02064.x
- Morey, R., Romeijn, J.-W., & Rouder, J.N. (2013). The humble Bayesian: model checking from a fully Bayesian perspective. *British Journal of Mathematical and Statistical Psychology*, *66*, 68–75. doi:10.1111/j.2044-8317.2011.02067.x
- Neal, R. (1996). *Bayesian learning for neural networks*. New York: Springer.
- Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, *324*, 81–85. doi:10.1126/science.1165893
- Senn, S. (2013). Comment on Gelman and Shalizi. *British Journal of Mathematical and Statistical Psychology*, *66*, 65–67. doi:10.1111/j.2044-8317.2011.02065.x

Received 8 June 2012