



PROJECT MUSE®

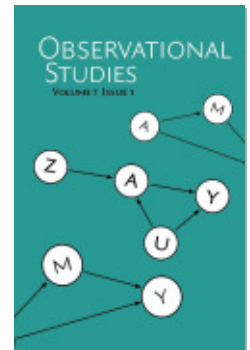
## Reflections on Breiman's Two Cultures of Statistical Modeling

Andrew Gelman

Observational Studies, Volume 7, Issue 1, 2021, pp. 95-98 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2021.0025>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/799750>

# Reflections on Breiman's Two Cultures of Statistical Modeling

**Andrew Gelman**

gelman@stat.columbia.edu

*Department of Statistics*

*Columbia University*

*New York, NY 10027, USA*

## Abstract

In his article on Two Cultures of Statistical Modeling, Leo Breiman argued for an algorithmic approach to statistics, as exemplified by his pathbreaking research on large regularized models that fit data and have good predictive properties but without attempting to capture true underlying structure. I think Breiman was right about the benefits of open-ended predictive methods for complex modern problems. I also discuss some points of disagreement, notably Breiman's dismissal of Bayesian methods, which I think reflected a misunderstanding on his part, in that he did not recognize that Bayesian inference can be viewed as regularized prediction and does not rely on an assumption that the fitted model is true. In retrospect, we can learn both from Breiman's deep foresight and from his occasional oversights.

**Keywords:** algorithms, Bayesian inference, prediction, statistical modeling

In an influential paper from 2001, the statistician Leo Breiman distinguished between two cultures in statistical modeling: "One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown." Breiman's "two cultures" article deserves its fame: it includes many interesting real-world examples and an empirical perspective which is a breath of fresh air compared to the usual standard approach of statistics papers at that time, which was a mix of definitions, theorems, and simulation studies showing the coverage of nominal 95% confidence intervals.

## Points of agreement

In his article, Breiman was capturing an important principle that I learned from Hal Stern: The most important thing is what data you use, not what you do with the data. A corollary to Stern's principle is that what makes a statistical method effective is that it facilitates the inclusion of more data.

A common feature of modern big-data approaches to statistics, including lasso, hierarchical Bayes, deep learning, and Breiman's own trees and forests, is regularization—estimating lots of parameters (or, equivalently, forming a complicated nonparametric prediction function) using some statistical tools to control overfitting, whether by the use of priors, penalty functions, cross-validation, or some mixture of these ideas. All these approaches to regu-

larization continue to be the topic of active research. I would not at all say that there is a unity among the competing classes of models—an additive model with interactions is not the same as a mixture of trees—but I take the fact that all these methods can, and do, solve real problems every day in settings where simple least squares and maximum likelihood would fail, as evidence of the benefit of regularization procedures that enable the fitting of complex response surfaces without immediate fear of overfitting. In recent years, even bastions of methodological conservatism such as econometrics have seen the benefits of moving beyond least squares and purportedly unbiased estimation. (I say “purportedly” because in practice applied inferences are often published conditional on statistical significance, a filter which leads to large “winner’s curse” or type M (magnitude) errors.)

Indeed, traditionally there has been a division of statistical thinking into two cultures in a way that is somewhat orthogonal to Breiman’s categories. The distinction I’m thinking of is between the complex modeling approach, in which we recognize that it is impossible to truly model the data so we construct a complex model or prediction algorithm so as to get as close as we can, or the reduced-form approach, in which we recognize that no model will be correct and so we try to produce an estimate that is interpretable as an aggregate linear fit or an average treatment effect in a way that is robust to inevitable model errors. Traditionally, the first approach is associated with statisticians and the second with econometricians. I say this division is orthogonal to Breiman’s because it strikes me that his two cultures both live within the complex modeling approach: nowhere, for example, is he suggesting to just perform least squares or compute some weighted average of the data. From this perspective, hierarchical Bayes and classification forests, different as they are both mathematically and in their computational implementation, are sisters under the skin, in that they are both examples of the fit-as-much-as-you-can-from-your-data attitude, rather than the go-for-robustness-using-a-reduced-form approach.

Reduced-form methods have not gone away, and there will always be a place for simple, robust methods, especially for clean designed experiments where there is less of a need to adjust for many potential biasing factors, but I think Breiman was right about the benefits of open-ended predictive methods for complex modern problems.

## Points of disagreement

I could fill several pages with what Breiman got right in his article and how his perspective and hugely important, outside-the-box contributions have led to progress in statistics, machine learning, and many areas of application. But I expect the other discussants will do this. So here I will focus on the places where I think he messed up, where his success developing and using certain statistical methods let him to mistakenly disparage other approaches.

Most notably, Breiman screwed up in his dismissal of Bayesian methods. In his 1997 article, “No Bayesians in foxholes,” he wrote, “I [Breiman] spent 13 years as a full-time consultant and continue to consult in many fields . . . Never once, either in my work with others or in anyone else’s published work in the fields in which I consulted, did I encounter the application of Bayesian methodology to real data. . . . All it would take to convince me [about Bayesian methods] are some major success stories in complex, high-dimensional problems where the Bayesian approach wins big compared to any frequentist approach. .

. . . A success story is a tough problem on which numbers of people have worked where a Bayesian approach has done demonstrably better than any other approach.” By that time, there were already many such success stories in fields ranging from psychometrics to political science to toxicology. He just (a) hadn’t encountered these successes personally, and (b) made no serious attempt to look for them. It’s ironic that Breiman wrote of the two cultures given the difficulty he had in understanding intellectual cultures other than his own.

Why does this matter? Can we take the good in Breiman’s article (its breadth of perspective, its predictive view of statistics) and the good in Breiman’s research (so many path-breaking methods that remain influential today) and set aside his misconceptions regarding Bayesian methods? Of course we can. What is important about a line of research is its successes, not its blind spots.

But it is instructive to look at these blind spots to help us going forward.

First, and most directly, the idea of hierarchical modeling is valuable, not just for its many applications in areas such as pharmacology and survey research, but for its more general relevance to problems of generalization. Just the other day, a colleague and I set up a hierarchical model for coronavirus tests, resolving a problem with a high-profile analysis that had inappropriately pooled data from several calibration studies. Hierarchical modeling directly allowed a partial-pooling compromise. We were under no illusion that our model was perfect; rather, it was a procedure that allowed us to follow Stern’s principle and include more information—in this case, the multiple different calibration studies, along with demographic and geographic information on the sample of people being tested. This sort of multilevel modeling is increasingly important in big-data settings such as opt-in surveys and messy observational studies in which we want to adjust for many factors that could bias our conclusions. As Breiman noted in his article, dimensionality should be a blessing rather than a curse if we move beyond naive optimization and include regularization and predictive model evaluation.

Second, Bayesian inference is central to many implementations of deep nets. Some of the best methods in machine learning use Bayesian inference as a way to average over uncertainty. A naive rejection of Bayesian data analysis would shut you out of some of the most effective tools out there. A safer approach would be to follow Brad Efron and be open to whatever works.

Third, it’s useful to recognize that even heroes have weaknesses. Following Breiman’s mistaken “no Bayesians in foxholes” claim, I coined the “foxhole fallacy,” which is the belief that there are no X’s in foxholes (where X = people who disagree with you on some issue of faith). When other people disagree with you, it can be a good idea to consider that they might be right! In this particular case, there are Bayesians in foxholes. When my colleagues and I analyze public opinion, we use Bayesian methods for real—and when the question is more important, we lean more heavily into the Bayesian paradigm. In saying this, I accept that other statisticians have other perspectives in their own foxholes. Much depends on what problems you are working on and what approaches you are already comfortable with. In a “foxhole” situation, it is typically rational to use the methods that have worked for you in the past. Recalling Stern’s principle, I hope that all of us when in foxholes will employ methods that can make use of as much information as possible.

## Looking forward

As noted above, Breiman’s insights into meta-statistics as well as his statistical ideas remain relevant today. His anti-Bayesianism was a blind spot but it should not have much effect on now we read his article on the two cultures, once we recognize that hierarchical Bayes can be viewed as an example of his preferred approach.

I will conclude by listing three ways that we can apply Breiman’s algorithmic perspective to current problems in statistics.

First is the problem of taking inference from existing data and generalizing to a new predictive problem. This is a central task of statistics and arises even with the cleanest experimental and survey data, as we are almost always interested in applying our findings to new cases and new scenarios. In such problems we should respect Breiman’s dictum that the data mechanism is unknown—indeed, in these problems of generalization there is typically no “data mechanism” at all. I think the concept of “the data mechanism” is itself a holdover from traditional theoretical statistics. We can apply hierarchical modeling to regularize our predictions and apply cross-validation to evaluate them, and new challenges arise when cross-validating structured rather than independently distributed data.

Second is the problem of model checking. Once we accept that our models are only approximations, this does not mean we can give up or relax into nihilism. By seeing how our models don’t fit the data, and where in data space this is happening, we can build better models and make better predictions. Models, being imperfect, are works in progress.

Third is the problem of black-box computing. Random forests, hierarchical Bayes, and deep learning all have in common that they can be difficult to understand (although, as Breiman notes, purportedly straightforward models such as logistic regression are not so easy to understand either, in practical settings with multiple predictors) and are fit by big computer programs that act for users as black boxes. Anyone who has worked with a black-box fitting algorithm will know the feeling of wanting to open up the box and improve the fit: these procedures often do this thing where they give the “wrong” answer, but it’s hard to guide the fit to where you want it to go. There is a tension between Stern’s principles of model checking, model improvement, and using more data, and the black-box feel of modern predictive models.

All this is to say that we have a lot more work to do, and seeing the strengths and weaknesses of Breiman’s arguments may point us in some helpful directions.

## Acknowledgments

We thank Dylan Small for helpful comments and the U.S. Office of Naval Research for financial support.