

# Entropy, Statistical Evidence, and Scientific Inference: Evidence Functions in Theory and Applications

Mark L. Taper<sup>1,\*</sup>, José Miguel Ponciano<sup>2,3</sup> and Brian Dennis<sup>4,5</sup>

<sup>1</sup> Department of Ecology, Montana State University, Bozeman, MT 59717, USA

<sup>2</sup> Biology Department, University of Florida, Gainesville, FL 32611, USA

<sup>3</sup> Mathematics Department, University of Florida, Gainesville, FL 32611, USA

<sup>4</sup> Department of Mathematics and Statistical Science, University of Idaho, Moscow, ID 83844, USA

<sup>5</sup> Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID 83844, USA

\* Correspondence: markltaper@gmail.com

**Scope and Goals of the Special Issue:** There is a growing realization that despite being the essential tool of modern data-based scientific discovery and model testing, statistics has major problems. The statistician Ionides speculated that most reported statistical results are false. The perceived failure of scientific studies to reproduce “statistically significant” results has scientists and statisticians alike wringing their hands and despairing of what has been named the “replication crisis”. Public confidence in science is weakening, and perhaps publicity about the replication crisis is contributing to this. Why have these problems in statistics remained uncorrected for so long? There has been a struggle of Wagnerian or titanic proportions between two schools of thought: classical statistics and Bayesian statistics. Neither has prevailed because each school has virtues, and each school has flaws. The virtues impel each school’s adherents to defend, while the flaws give both schools ample opportunity to attack the other. This impasse has been dubbed by the philosopher Deborah Mayo [1] the “Statistics Wars”. In our view, the way forward, the path between the paradigms, is evidential statistics. For many years, there has been a growing understanding of the importance of theory in Ecology and Evolution, with theory being an organized complex of models [2–5]. Thus, a more model-centric approach to statistics seems appropriate. Practically, science functions by comparing the relative fit of models to data. Science advances when existing (possibly good) models are replaced with new (hopefully better) models. Modern statistical evidence compares the relative support in scientific data for mathematical models. The fundamental tool of model comparison is the evidence function.

In all of the relevant literatures (philosophical, statistical, and scientific) there is frequent equivocation between the terms (and concepts) of “model” and “hypothesis”. In our usage, a model is a mechanism to generate data or pseudo data. A model may be mathematical, algorithmic, or even physical. A hypothesis, on the other hand, is a statement that may be either true or false. Philosophers, statisticians, and scientists often use the two terms interchangeably. We find this confusing. This is not to say that individual writers are necessarily confused. For example, the statistician Richard Royall in his ground breaking 1997 volume [6] is quite clear when he says on page 4 that “A simple statistical hypothesis is one that complexly specifies the probability distribution of an observable random variable.” He is describing what we would speak of as a model. However, for every model, there a host of associated hypotheses such as: the model is true, the model describes reality within a specified level of accuracy, the best value for a particular parameter fall within a specified interval, and so forth.

Royall [7] has said “Statistics today is in a conceptual and theoretical mess.” He noted [7,8] that there are three generic questions regarding models or hypotheses after data have been collected:



**Citation:** Taper, M.L.; Ponciano, J.M.; Dennis, B. Entropy, Statistical Evidence, and Scientific Inference: Evidence Functions in Theory and Applications. *Entropy* **2022**, *24*, 1273. <https://doi.org/10.3390/e24091273>

Received: 25 August 2022

Accepted: 26 August 2022

Published: 9 September 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. What do I believe, now that I have this observation?
2. What should I do, now that I have this observation?
3. What does this observation tell me about A versus B? (How should I interpret this observation as evidence regarding A versus B?) (Royall, 1997 page 4).

Royall [6] identified the crux of problem in statistics as a failure of both the classical and Bayesian schools to address head-on a question fundamental to science: “when does a given set of observations support one statistical hypothesis over another?” In his 1997 monograph [6], Royall sought to take a stream of thought running from Venn [9] to Barnard [8] to Birnbaum [10] to Hacking [11] and finally to Edwards [12] elevate it to a river of a paradigm. Royall [6] consciously tried to devise a system that retained virtues of the contesting schools while jettisoning flaws. Royall axiomatically based his evidential statistics on the law of likelihood [11] and the likelihood principle (LP) [10] and utilized the likelihood ratio (LR) as the canonical measure of statistical evidence. This “leads to statistical methods that have a compelling rationale and elegant simplicity” [6].

Unfortunately, the likelihood principle has been strongly critiqued by many. These criticisms fall roughly into three categories: (1) Questions about the legitimacy of Birnbaum’s 1962 derivation (e.g., Mayo [13], but see Gardenberger [14]). (2) Examples where application of the LP leads to paradoxical results (e.g., Cornfield [15]). Additionally, (3) questions about the LP applicability to scientific inference [13,16–19], given that it was formulated and proved under a perfect model specification scenario, a criterion that is rarely met, if at all, in day-to-day scientific modeling settings.

The greatest problem with the LP is that it is an incredibly limited statement: It only says that if the generating process is exactly given by a known parametric model, then all the information for comparing parameter values is contained in their likelihoods. The principle does not address many things necessary for the prosecution of science. The LP is silent regarding the comparison of different model forms, it is silent as to what happens when the model is misspecified, and it is silent on the quantification of error. Consequently, the LP has been challenged as a foundational tool for scientific inference ever since it was named [20]:

*In advocating the likelihood principle as a basic proposition in the theory of inference, it is necessary to emphasize that it is not proposed that it should be used uncritically unless the model is known very precisely. [...] If the model has to be guessed, then the nature of the inference becomes much less precise than is suggested by the formal statement of the likelihood principle. (Page 323, Barnard et al. 1962)*

We argue that, because scientists are not omniscient, the model virtually always must be “guessed” to some degree. Consequently, the most useful way to think about the LP is as an instructive limiting case—an ideal gas law of sorts. It gives us ideas as to how to fruitfully pursue inference in more realistic cases.

The LR works well as an evidence measure when comparing two statistical models with no unknown parameters under the correct model specification. Such models are known in statistics as “simple hypotheses.” Difficulties arise when model misspecification, model selection, and unknown parameters (including “nuisance parameters”) are involved. In 2004 the statistician Subhash Lele [21] recognized that the log of the likelihood ratio,  $\log LR$ , between two simple models is an estimate of the contrast of Kullback–Leibler divergences of each of the two models to the underlying generating process. From this realization, he developed the concept of evidence functions as a generalization of the  $\log LR$ . Evidence functions are essentially contrasts of generalized entropy discrepancies. Corrections can often be constructed, substituting measures that look like the LR or  $\log LR$ , but are not based entirely on pure likelihood ratios [22–25].

We feel that evidence functions have intuitive appeal and compelling rationale—exceeding even that of the LR. Statistical evidence for one model over another model is just a data-based estimate of the difference in discrepancies of each of two models to the unknown underlying generative process. Furthermore, it is suggested that evidence functions should

be developed in accordance with a list of desiderata, each chosen to confer good inferential properties to evidence [21,26,27]. We feel that before proceeding to questions 1 and/or 2, an analyst should endeavor to first answer the evidence question, that is to “let the data speak for themselves” (This quote is often cited as stemming from R. J. Fisher, but we have not been able to verify the attribution). The desiderata in Box 1 are constructed to promote this goal in as clean a statistical fashion through the construction of sound evidence functions.

**Box 1.** Desiderata for evidence quoted, with minor emendations, from (Taper and Ponciano [26]).

- D1. Evidence should be a data-based estimate of a contrast of the divergences of each of two models from the data generating process.
- D2. Evidence should be a continuous function of data. This means that there is no threshold that must be passed before something is counted as evidence.
- D3. The reliability of evidential statements should be quantifiable.
- D4. Evidence should be public not private or personal.
- D5. Evidence should be portable, that is it should be transferable from person to person.
- D6. Evidence should be accumulable: If two data sets relate the same pair of models, then the evidence should be combinable in some fashion, and any evidence collected should bear on any future inferences regarding the models in question.
- D7. Evidence should not depend on the personal idiosyncrasies of model formulation. By this we mean that evidence functions should be both scale and transformation invariant.
- D8. Consistency, evidence for the true model/parameter is maximized at the true value only if the true model is in the model set, or at the best projection into the model set if it is not.

Using data, evidence functions *estimate* the relative divergence between models and the generating process in some specific metric. Thus, evidence functions can be defined using divergences other than the KL to deal with specific kinds of models, data, or types of questions being asked [21,28]. For some of these cases, such as the important question of choosing between parametric model families, the likelihood principle cannot hold by definition, at least not without modification such as computing differences of expected log-likelihood as suggested by Akaike [29].

This generalization of the LR to evidence functions does not represent any loss of capacity as the LR and all of the substitute measures mentioned above can now be seen as special cases of evidence functions. The log likelihood ratio still holds a place of pride. Lele [21] observed that evidence functions based on other divergences have a lower probability of correct model identifications at every given sample size—under correct model specification.

Royall [6] made at least two critical advances: First, was the elimination of the null model/alternative model distinction, conceptually and statistically. Both models are treated equally and symmetrically. Additionally, second, the accept/reject dichotomy is replaced with a trichotomy. Possible outcomes are strong evidence for model 1, weak and inconclusive evidence, and strong evidence for model 2. Taper et al. [19] suggest further dividing the outcome continuum into strong evidence for model 1, prognostic evidence for model 1, weak evidence not really favoring either model, prognostic evidence for model 2, and strong evidence for model 2. Prognostic evidence is suggestive of a model, but not sufficiently strong to stand on its own without collateral evidence.

Since then, modern evidential statistics has extended the concept of evidence to statistical cases with unknown parameters and misspecified models, but it realizes that a frequentist error structure can be created for both pre-data design and post-data inference [15,19,23,30]. Further, both conditional uncertainty (i.e., uncertainty in evidence given the data in hand) and unconditional uncertainty (i.e., uncertainty in evidence over resampling of data) can be calculated [19,31].

Testing in classical statistics only conceives of the strength of evidence against a null hypothesis as an error probability, either significance level or size of test. There are several substantial improvements in the handling of errors in evidential statistics from that in

classical statistics. First, the concepts of evidence and frequentist error are now distinct, and their inferential impacts can be considered separately [30,32,33]. Second, the asymptotic error structures for evidence, briefly explained below, are vastly superior to those found in Neyman–Pearson hypothesis testing, NPHT.

It is certainly desirable that inferences should get better as the sample size,  $n$ , increases. This is not the case for NPHT—even under correct model specification. Inferences are bound to error rates, and Type I error rates ( $\alpha$ ) are constant over all sample sizes. In evidence, the error rates are the probabilities: (a) of misleading evidence,  $M_i$ , and (b) of weak evidence,  $W$ . The quantities  $M_1$  and  $M_2$  for two models = are roughly analogous to the Type I and Type II error rates  $\alpha$  and  $\beta$  in NPHT. In an evidential analysis, under correct model specification, both  $M_1$  and  $M_2$ , along with  $W$ , go to zero as sample size increases, in strong contrast to the error structure of NPHT. Further, the evidential approach allows an experimenter to control the error by setting a priori the level of evidence,  $k$ , that will be considered strong [6,23]. Presetting  $k$  is analogous to setting  $\alpha$  before conducting an experiment to be analyzed using NPHT.

Model misspecification is deleterious to the error control of both NPHT and evidential model comparisons, but the impacts are much worse for NPHT than for evidence. NPHT realized error rates can be less than, equal to or greater than the nominal size,  $\alpha$ , depending on the nature of misspecification, which is disastrous for a procedure that relies entirely on its nominal error rate for inference. To make things even worse, in some realistic situations, realized error rates will actually increase with increasing sample size—eventually reaching a certainty of error [30]. Interestingly, the geometry of the model space is estimable [34] and influences the model identification error probabilities [30]. Further, if nonparametric estimates of data entropy are available then absolute distances from the generating process to models are calculable, not just relative distances [34]. See [19,30,34] for detailed discussions of the impact of model misspecification on evidential assessment.

Under misspecification, the maximum error rates for evidential model identification can be as large as 50%. This occurs when two distinct models are equally divergent from the generating process, an intuitive result. However, for all model comparisons except for this limiting case, both evidential error rates ( $M$  and  $W$ ) decline to zero as sample size increases.

Another interesting tool for assessing uncertainty in an evidential analysis is the confidence interval for your evidence. Evidence is for comparing models; your evidence is a point estimate of the differences of the divergences each model from the true data generating process. *Even when models are mis-specified*, interval estimates for evidence with valid coverage properties can be produced whenever bootstrapping is possible [19].

Thus far, we have compared the evidential approach to statistics with the classical approach, but what about Bayesianism? Bayesian statistical thinking is not a monolith, but a complex stew of shading opinions. I. J. Good [35] waggishly argued that there are at least 46,657 different kinds of potential Bayesians. In seeking some principle common to all he said, “All Bayesians, as I understand the term, believe that it is usually meaningful to talk about the probability of a hypothesis”. It follows from this characterization that Bayesians fundamentally do not address the evidence question because the evidence question is always comparative, see Royall’s question 3 above. For extensive discussion of the confirmation/evidence distinction see [36,37]. While seeking to confirm one’s beliefs is a natural tendency, it runs the risk of being smacked in the head by reality [8]:

*The advantages of odds are even more striking in relation to hypotheses. To speak of the probability of a hypothesis implies the possibility of an exhaustive enumeration of all possible hypotheses, which implies a degree of rigidity foreign to the true scientific spirit. We should always admit the possibility that our experimental results may be best accounted for by a hypothesis which never entered our own heads. (Barnard, 1949 page 136)*

We can broadly categorize Bayesians into three classes: subjective Bayesians, objective Bayesians, and empirical Bayesians. Further we feel that the empirical Bayes is substantively different from the other two and only Bayes by a historical fluke of nomenclature, as the

method could have been just as accurately described as a mixture model or semiparametric likelihoodism if the prior is estimated nonparametrically.

The distinction between empirical Bayes on one hand and subjective and objective Bayes on the other hand is that the results of an empirical Bayes analysis are invariant to the transformation of the parameters while analyses using subjective or objective Bayes are not. This lack of invariance is not just a philosophical problem but has real world consequences [38].

There is a tendency among scientists to use “noninformative priors”, perhaps because personal idiosyncrasy of subjective priors seems somehow inappropriate for science. We applaud this impulse but point out that the solution is apocryphal. A corollary to the problem of transformation dependence in Bayesian analysis is that noninformative priors do not really exist except for categorical models. Every noninformative prior is informative for another transformation of the parameters and, conversely, every informative prior is non-informative on an appropriately transformed parameter space [38]. Even more cryptically, the induced priors on functions of parameters are generally, informative even though priors on the original parameter space are non-informative.

Another difficulty with Bayesian inference is that parameter non-identifiability is masked in the posterior distribution. It can happen for certain situations of model and data that the likelihood surface is flat, or nearly so, in a particular region of the parameters. Thus, the data give no, or little, information on parameter values so the posterior is entirely, or mostly, dependent on the prior [38]. This is particularly a problem for complex situations where non-identifiability may not be easily recognizable [39,40].

It has been repeatedly pointed out that the posteriors of any Bayesian analysis (subjective or objective, but excluding empirical) is ineluctably a distribution of belief [16,26,41]. No amount of data changes this fact. To many, this seems a philosophical nicety that is completely overwhelmed by the ability to provide estimates for difficult problems. Latent variable and other hierarchical models can be particularly challenging to a likelihood approach because their optimization requires repeated calculation of multiple integrals of very high order [42,43]. Monte Carlo Markov chain methods allow the estimation of the posterior distributions of complicated Bayesian problems through simulation [42,43]. A great deal of mileage has been made of the Walker [44] theorem, which proves that for i.i.d. data a Bayesian posterior distribution of an identifiable parameter vector  $\theta$  converges to the sampling distribution of  $\hat{\theta}$  (the maximum likelihood estimate) as sample size goes to infinity (e.g., Yamamura [45]).

Unfortunately, real-world inference is not asymptotic. While simple models often achieve near-asymptotic behavior with quite small sample sizes, model complexities and prior infelicities can make the approach to asymptotic behavior quite slow. There is no way of knowing if a Bayesian analysis has converged to its large sample behavior except to data clone [46,47]. Data cloning is an algorithmic device to eliminate the influence of the prior from a Bayesian estimation. Data cloning provides a full range of frequentist/evidential statistical tools including maximum likelihood estimation, hypothesis testing, confidence intervals (Wald and profile), model identification via information criteria, and diagnostics for parameter identifiability [48–51].

A Bayesian analysis requires that the prior belief distribution integrates to 1. Thus, fundamentally, but implicitly, a Bayesian analysis assumes that the generating model is in the model set. When the generating process is not in the model set, a Bayesian analysis will be overconfident [52]. We agree with Barnard [8] (see quote above) and Box [53,54], who asserted repeatedly that “all models are wrong” and argue strongly that the generating process is almost never in any real scientific model set. Some Bayesians [55] address this problem by placing priors on models; this creates a plausible personal probability. However, this means that model probabilities are conditional on the model set. Bayesian scientists are faced with the dilemma of either claiming to be omniscient or accepting that the overarching idea of a probability for a model, that Good [35] (see quote above) felt was the common glue that held all Bayesians together, has shrunk so small that it can sleep in a



teacup. All Bayesian probabilities, all Bayesian confirmations are only consistent with the real world (that is a world with model misspecification) as comparative concepts [37,56]. Some Bayesian statisticians and scientists fruitfully acknowledge these limitations and see Bayesian computation as an estimation device for complex models that needs to be followed by non-Bayesian calibration, confidence, and model-checking [57].

The aim of this special feature in the journal *Entropy* is to continue the development and dissemination of evidential statistics, building on the recent research topic in *Frontiers in Ecology and Evolution* (<https://www.frontiersin.org/research-topics/7282/evidential-statistics-model-identification-and-science> accessed 25 August 2022). We believe that evidential statistics presents an opportunity for increased clarity in scientific inference. The evidential project draws strengths from both of the opposing camps in the “Statistics Wars” while simultaneously rejecting some of the flaws of each. The evidential approach promotes efficient scientific progress by encouraging more active (yet controlled) exploration of model space [26,34]. We also believe that widespread adoption of evidential practices should act to alleviate many of the deleterious social behaviors of the scientific community that are partial causes of the replication crisis. These include “cherry picking”, “HARKing”, “the file drawer problem”, and of course misunderstanding and/or misrepresentation of the kind of uncertainty being reported [58].

We are soliciting papers that convey basic concepts and papers that convey technical subtleties sufficient to conduct real scientific research, as well as practical advice that can be incorporated into the teaching of undergraduate and graduate courses. Additionally, this special feature would benefit from application to science beyond simple examples imbedded within formulations, proofs, computer code, or logical arguments. Specifically, we are seeking papers that use the evidentialist approach to address pressing scientific questions or demonstrate the strengths and limitations of the evidentialist approach in scientific research [26,59].

The topic will consist of a mix of new original research, reviews, mini-reviews, opinions, and commentaries and perspectives on topics related to evidential statistics. New statistical work is encouraged, nevertheless, all papers will need to spend significant effort to explain goals, utility, and application of methods to working scientists. To further this goal, collaboration among statisticians and more empirical scientists is also encouraged. In the interest of sharpening the discussion, papers by Bayesians, likelihoodists, NP testers, severe testers, and others that explicitly critique the modern evidential project and any or all points touched on above are also welcome.

**Author Contributions:** All authors jointly conceived of the project, M.L.T. wrote the initial draft, all authors contributed to manuscript revision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This essay has been substantially improved by the critical reading and insightful comments of Prasanta S. Bandyopadhyay, Gordon Brittan, Christopher Jerde, and Subhash R. Lele.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mayo, D.G. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*; Cambridge University Press: Cambridge, UK, 2018.
2. Scheiner, S.M. Toward a conceptual framework for biology. *Q. Rev. Biol.* **2010**, *85*, 293–318. [[CrossRef](#)] [[PubMed](#)]
3. Scheiner, S.M.; Holt, R.D. Evidential Statistics in Model and Theory Development. *Front. Ecol. Evol.* **2019**, *7*, 306. [[CrossRef](#)]
4. Scheiner, S.M.; Willig, M.R. A general theory of ecology. *Theor. Ecol.* **2008**, *1*, 21–28. [[CrossRef](#)]
5. Zamer, W.E.; Scheiner, S.M. A Conceptual Framework for Organismal Biology: Linking Theories, Models, and Data. *Integr. Comp. Biol.* **2014**, *54*, 736–756. [[CrossRef](#)] [[PubMed](#)]
6. Royall, R.M. *Statistical Evidence: A Likelihood Paradigm*; Chapman & Hall: London, UK, 1997.
7. Royall, R.M. The Likelihood Paradigm for Statistical Evidence. In *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations*; Taper, M.L., Lele, S.R., Eds.; The University of Chicago Press: Chicago, IL, USA, 2004; pp. 119–152.

8. Barnard, G.A. *Statistical Inference*. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **1949**, *11*, 115–149. [[CrossRef](#)]
9. Venn, J. *The Logic of Chance*, 2nd ed.; Reprinted 1962; Chelsea Publishing Co.: New York, NY, USA, 1876.
10. Birnbaum, A. On the foundations of statistical inference. *J. Am. Stat. Assoc.* **1962**, *57*, 269–306. [[CrossRef](#)]
11. Hacking, I. *Logic of Statistical Inference*; Cambridge University Press: Cambridge, UK, 1965.
12. Edwards, A.W.F. *Likelihood*; Cambridge University Press: Cambridge, UK, 1972.
13. Mayo, D.G. On the Birnbaum Argument for the Strong Likelihood Principle. *Stat. Sci.* **2014**, *29*, 227–239. [[CrossRef](#)]
14. Gandenberger, G. A new proof of the likelihood principle. *Br. J. Philos. Sci.* **2015**, *66*, 475–503. [[CrossRef](#)]
15. Cornfield, J. Sequential trials, sequential analysis and likelihood principle. *Am. Stat.* **1966**, *20*, 18–23.
16. Dennis, B. Discussion: Should Ecologists Become Bayesians? *Ecol. Appl.* **1996**, *6*, 1095–1103. [[CrossRef](#)]
17. Sprott, D.A. *Statistical Inference in Science*; Springer: New York, NY, USA, 2000.
18. Cox, D.R. Commentary on The Likelihood Paradigm for Statistical Evidence by R. Royall. In *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations*; Taper, M.L., Lele, S.R., Eds.; University of Chicago Press: Chicago, IL, USA, 2004; pp. 119–152.
19. Taper, M.L.; Lele, S.R.; Ponciano, J.M.; Dennis, B.; Jerde, C.L. Assessing the Global and Local Uncertainty of Scientific Evidence in the Presence of Model Misspecification. *Front. Ecol. Evol.* **2021**, *9*, 679155. [[CrossRef](#)]
20. Barnard, G.A.; Jenkins, G.M.; Winsten, C.B. Likelihood Inference and Time-Series. *J. R. Stat. Soc. Ser. Gen.* **1962**, *125*, 321–372. [[CrossRef](#)]
21. Lele, S.R. Evidence functions and the optimality of the law of likelihood. In *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations*; Taper, M.L., Lele, S.R., Eds.; The University of Chicago Press: Chicago, IL, USA, 2004.
22. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
23. Royall, R.M. On the Probability of Observing Misleading Statistical Evidence. *J. Am. Stat. Assoc.* **2000**, *95*, 760–780. [[CrossRef](#)]
24. Royall, R.; Tsou, T.S. Interpreting statistical evidence by using imperfect models: Robust adjusted likelihood functions. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **2003**, *65*, 391–404. [[CrossRef](#)]
25. Baskurt, Z.; Strug, L.J. Genetic association analysis with pedigrees: Direct inference using the composite likelihood ratio. *Genet. Epidemiol.* **2018**, *42*, 826–837. [[CrossRef](#)]
26. Taper, M.L.; Ponciano, J.M. Evidential statistics as a statistical modern synthesis to support 21st century science. *Popul. Ecol.* **2016**, *58*, 9–29. [[CrossRef](#)]
27. Jerde, C.L.; Kraskura, K.; Eliason, E.J.; Csik, S.R.; Stier, A.C.; Taper, M.L. Strong Evidence for an Intraspecific Metabolic Scaling Coefficient Near 0.89 in Fish. *Front. Physiol.* **2019**, *10*, 1166. [[CrossRef](#)]
28. Markatou, M.; Sofikitou, E.M. Statistical Distances and the Construction of Evidence Functions for Model Adequacy. *Front. Ecol. Evol.* **2019**, *7*, 447. [[CrossRef](#)]
29. Akaike, H. Information Theory as an Extension of the Maximum Likelihood Principle. In *Second International Symposium on Information Theory*; Akademiai Kiado: Budapest, Hungary, 1973; pp. 267–281.
30. Dennis, B.; Ponciano, J.M.; Taper, M.L.; Lele, S.R. Errors in Statistical Inference Under Model Misspecification: Evidence, Hypothesis Testing, and AIC. *Front. Ecol. Evol.* **2019**, *7*, 372. [[CrossRef](#)]
31. Lele, S.R. How Should We Quantify Uncertainty in Statistical Inference? *Front. Ecol. Evol.* **2020**, *8*, 35. [[CrossRef](#)]
32. Taper, M.L.; Lele, S.R. Evidence, evidence functions, and error probabilities. In *Philosophy of Statistics*; Bandyopadhyay, S., Forster, M., Eds.; Elsevier: Amsterdam, The Netherlands, 2011; pp. 513–532.
33. Strug, L.J.; Hodge, S.E. An alternative foundation for the planning and evaluation of linkage analysis I. Decoupling ‘error probabilities’ from ‘measures of evidence’. *Hum. Hered.* **2006**, *61*, 166–188. [[CrossRef](#)] [[PubMed](#)]
34. Ponciano, J.M.; Taper, M.L. Model Projections in Model Space: A Geometric Interpretation of the AIC Allows Estimating the Distance Between Truth and Approximating Models. *Front. Ecol. Evol.* **2019**, *7*, 413. [[CrossRef](#)] [[PubMed](#)]
35. Good, I.J. 46656 Varieties of Bayesians. *Am. Stat.* **1971**, *25*, 62.
36. Bandyopadhyay, P.S.; Brittan, G., Jr.; Taper, M.L. *Belief, Evidence, and Uncertainty: Problems of Epistemic Inference*; Springer: Cham, Switzerland, 2016.
37. Bandyopadhyay, P.S.; Taper, M.L.; Brittan, G., Jr. Non-Bayesian Accounts of Evidence: Howson’s Counterexample Countered. *Int. Stud. Philos. Sci.* **2017**, *30*, 291–298. [[CrossRef](#)]
38. Lele, S.R. Consequences of lack of parameterization invariance of non-informative Bayesian analysis for wildlife management: Survival of San Joaquin kit fox and declines in amphibian populations. *Front. Ecol. Evol.* **2020**, *7*, 501. [[CrossRef](#)]
39. Gustafson, P. On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. *Stat. Sci.* **2005**, *20*, 111–129. [[CrossRef](#)]
40. Lele, S.R. Model complexity and information in the data: Could it be a house built on sand? *Ecology* **2010**, *91*, 3493–3496. [[CrossRef](#)]
41. Brittan, G.J.; Bandyopadhyay, P.S. Ecology, Evidence, and Objectivity: In Search of a Bias-Free Methodology. *Front. Ecol. Evol.* **2019**, *7*, 399. [[CrossRef](#)]
42. Robert, C.; Casella, G. *Monte Carlo Statistical Methods*, 2nd ed.; Springer: New York, NY, USA, 2004.
43. Robert, C.; Casella, G. A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Stat. Sci.* **2011**, *26*, 102–115. [[CrossRef](#)]
44. Walker, A.M. On asymptotic behaviour of posterior distributions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1969**, *31*, 80–88. [[CrossRef](#)]

45. Yamamura, K. Bayes estimates as an approximation to maximum likelihood estimates. *Popul. Ecol.* **2016**, *58*, 45–52. [[CrossRef](#)]
46. Lele, S.R.; Dennis, B.; Lutscher, F. Data cloning: Easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecol. Lett.* **2007**, *10*, 551–563. [[CrossRef](#)]
47. Jacquier, E.; Johannes, M.; Polson, N. MCMC maximum likelihood for latent state models. *J. Econom.* **2007**, *137*, 615–640. [[CrossRef](#)]
48. Ponciano, J.M.; Taper, M.L.; Dennis, B.; Lele, S.R. Hierarchical models in ecology: Confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology* **2009**, *90*, 356–362. [[CrossRef](#)]
49. Lele, S.R.; Nadeem, K.; Schmuland, B. Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning. *J. Am. Stat. Assoc.* **2010**, *105*, 1617–1625. [[CrossRef](#)]
50. Ponciano, J.M.; Burleigh, J.G.; Braun, E.L.; Taper, M.L. Assessing Parameter Identifiability in Phylogenetic Models Using Data Cloning. *Syst. Biol.* **2012**, *61*, 955–972. [[CrossRef](#)]
51. Campbell, D.; Lele, S. An ANOVA test for parameter estimability using data cloning with application to statistical inference for dynamic systems. *Comput. Stat. Data Anal.* **2014**, *70*, 257–267. [[CrossRef](#)]
52. Yang, Z.; Zhu, T. Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 1854–1859. [[CrossRef](#)]
53. Box, G.E.P. Science and Statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799. [[CrossRef](#)]
54. Box, G.E.P. Robustness in the strategy of scientific model building. In *Robustness in Statistics*; Launer, R.L., Wilkinson, G.N., Eds.; Academic Press: New York, NY, USA, 1979; pp. 201–236.
55. Lindley, D.V. The philosophy of statistics. *J. R. Stat. Soc. Ser. D-Stat.* **2000**, *49*, 293–319. [[CrossRef](#)]
56. Howson, C. Exhuming the No-Miracles Argument. *Analysis* **2013**, *73*, 205–211. [[CrossRef](#)]
57. Gelman, A.; Shalizi, C.R. Philosophy and the practice of Bayesian statistics. *Br. J. Math. Stat. Psychol.* **2013**, *66*, 8–38. [[CrossRef](#)] [[PubMed](#)]
58. Taper, M.L.; Ponciano, J.M.; Toquenaga, Y. Editorial: Evidential Statistics, Model Identification, and Science. *Front. Ecol. Evol.* **2022**, *10*, 883456. [[CrossRef](#)]
59. Toquenaga, Y.; Gagné, T. The Evidential Statistics of Genetic Assembly: Bootstrapping a Reference Sequence. *Front. Ecol. Evol.* **2021**, *9*, 614374. [[CrossRef](#)]