Methods in Psychology 2 (2020) 100020

Contents lists available at ScienceDirect

Methods in Psychology

journal homepage: www.journals.elsevier.com/methods-in-psychology

What can psychology's statistics reformers learn from the error-statistical perspective? ☆

Brian D. Haig

University of Canterbury, New Zealand

ARTICLE INFO

Keywords:
The error-statistical perspective
The new statistics
Bayesian statistics
Falsificationism
Hierarchy of models
Philosophy of statistics

ABSTRACT

In this article, I critically evaluate two major contemporary proposals for reforming statistical thinking in psychology: The recommendation that psychology should employ the "new statistics" in its research practice, and the alternative proposal that it should embrace Bayesian statistics. I do this from the vantage point of the modern error-statistical perspective, which emphasizes the importance of the severe testing of knowledge claims. I also show how this error-statistical perspective improves our understanding of the nature of science by adopting a workable process of falsification and by structuring inquiry in terms of a hierarchy of models. Before concluding, I briefly discuss the importance of the philosophy of statistics for improving our understanding of statistical thinking.

1. Introduction

Psychology has been prominent among a number of disciplines that have proposed statistical reforms for improving our understanding and use of statistics in research. However, despite being at the forefront of these reforms, psychology has ignored the philosophy of statistics to its detriment. In this article, I consider, in a broad-brush way, two major proposals that feature prominently in psychology's current methodological reform literature: The recommendation that psychology should employ the so-called "new statistics" in its research practice, and the alternative proposal that psychology should embrace Bayesian statistics. I evaluate each from the vantage point of the error-statistical philosophy, which, I believe, is the most coherent perspective on statistics available to us. Before concluding, I discuss two interesting features of the conception of science adopted by the error-statistical perspective, along with brief remarks about the value of the philosophy of statistics for deepening our understanding of statistics.

2. The error-statistical perspective

The error-statistical perspective employed in this article is that of Deborah Mayo, sometimes in collaboration with Aris Spanos (Mayo, 1996, 2018; Mayo & Spanos, 2010, 2011). This perspective is landmarked by two major works. The first is Mayo's ground-breaking book, *Error and the growth of statistical knowledge* (1996), which presented the first extensive formulation of her error-statistical perspective on statistical inference. This philosophy provides a systematic understanding of experimental reasoning in science that uses frequentist statistics in order to manage error. Hence, its name. The novelty of the book lay in the fact that it employed ideas in statistical science to shed light on philosophical problems to do with the nature of evidence and inference.

The second book is Mayo's recently published *Statistical inference as severe testing* (2018). In contrast with the first book, this work focuses on problems arising from statistical practice, but endeavors to solve them by probing their foundations from the related vantage points of the philosophy of science and the philosophy of statistics. By dealing with the vexed problems of current statistical practice, this book is a valuable repository of ideas, insights, and solutions designed to help a broad readership deal with the current crisis in statistics. Because my focus is on statistical reforms in psychology, I draw mainly from the resources contained in the second book.

Fundamental disputes about the nature and foundations of statistical inference are long-standing and ongoing. Most prominent have been the numerous debates between, and within, frequentist and Bayesian camps. Cutting across these debates have been more recent attempts to unify and reconcile rival outlooks, which have complexified the statistical

☆ This article is based on an invited commentary on Deborah Mayo's book, *Statistical inference as severe testing*: *How to get beyond the statistics wars* (Cambridge University Press, 2018), which appeared at https://statmodeling.stat.colombia.edu/2019/04/12 It is adapted with permission. I thank Mayo for helpful feedback on an earlier draft.

E-mail address: brian.haig@canterbury.ac.nz.

https://doi.org/10.1016/j.metip.2020.100020
Received 15 December 2019; Received in revised form 22 March 2020; Accepted 23 March 2020
Available online 29 March 2020

landscape. Today, these endeavors fuel the ongoing concern that psychology and many sciences have with replication failures, questionable research practices, and the strong demand for an improvement of research integrity. Mayo refers to debates about these concerns as the "statistics wars". With the addition of *Statistical inference as severe testing* to the error-statistical corpus, it is fair to say that the error-statistical outlook now has the resources to enable statisticians and scientists to understand and advance beyond the bounds of these statistics wars.

The strengths of the error-statistical approach are considerable (Haig, 2017; Spanos, 2019a, 2019b), and I believe that they combine to give us the most coherent philosophy of statistics currently available. For the purpose of this article, it suffices to say that the error-statistical approach contains the methodological and conceptual resources that enable one to diagnose and overcome the common misunderstandings of widely used frequentist statistical methods such as tests of significance. It also provides a trenchant critique of Bayesian ways of thinking in statistics. I will draw from these two strands of the error-statistical perspective to inform my critical evaluation of the new statistics and the Bayesian alternative.

Because the error-statistical and Bayesian outlooks are so different, some might consider it unfair to use the former to critique the latter. My response to this worry is three-fold: First, perspective-taking is an unavoidable feature of the human condition; we cannot rise above our human conceptual frameworks and adopt a position from nowhere. Second, in thinking things through, we often find it useful to proceed by contrast, rather than direct analysis. Indeed, the error-statistical outlook on statistics was originally developed in part by using the Bayesian outlook as a foil. And third, strong debates between Bayesians and frequentists have a long history, and they have helped shape the character of these two alternative outlooks on statistics. By participating in these debates, the error-statistical perspective is itself unavoidably controversial.

3. The new statistics

For decades, numerous calls have been made for replacing tests of statistical significance with alternative statistical methods. The new statistics, which urges the abandonment of null hypothesis significance testing (NHST), and the adoption of effect sizes, confidence intervals, and meta-analysis as a replacement package, is one such reform movement (Calin-Jageman and Cumming, 2019; Cumming, 2012, 2014). It has been heavily promoted in psychological circles and touted as a much-needed successor to NHST, which is deemed to be broken-backed. *Psychological Science*, which is the flagship journal of the Association for Psychological Science, endorsed the use of the new statistics, wherever appropriate (Eich, 2014). In fact, the new statistics might be considered the Association's current quasi-official position on statistical inference. Although the error-statistical outlook does not directly address the new statistics movement, its suggestions for overcoming the statistics wars contain insights about statistics that can be employed to mount a powerful challenge to the integrity of that movement.

3.1. Null hypothesis significance testing

The new statisticians contend that NHST has major flaws and recommend replacing it with their favored statistical methods. Prominent among the flaws are the familiar claims that NHST encourages dichotomous thinking, and that it comprises an indefensible amalgam of the Fisherian and Neyman-Pearson schools of thought. However, neither of these features applies to the error-statistical understanding of NHST. The claim that we should abandon NHST because it leads to dichotomous thinking is unconvincing because it is leveled at the misuse of a statistical test that arises from its mechanical application and a poor understanding of its foundations. By contrast, the error-statistical perspective advocates the flexible use of levels of significance tailored to the case at hand as well as reporting of exact p values – a position that Fisher himself came to hold.

Further, the error-statistical perspective makes clear that the common understanding of the amalgam that is NHST is not an amalgam of Fisher's and Neyman and Pearson's thinking on the matter, especially their mature thought. Further, the error-statistical outlook can accommodate both evidential and behavioural interpretations of NHST, respectively serving *probative* and *performance* goals, to use Mayo's suggestive terms. The error-statistical perspective urges us to move beyond the claim that NHST is an inchoate hybrid. Based on a close reading of the historical record, Mayo argues that Fisher and Neyman and Pearson should be interpreted as compatibilists, and that focusing on the vitriolic exchanges between Fisher and Neyman prevents one from seeing how their views dovetail. Importantly, Mayo formulates the error-statistical perspective on NHST by assembling insights from these founding fathers, and additional sources, into a coherent hybrid. There is much to be said for replacing psychology's fixation on the muddle that is NHST with the error-statistical perspective on significance testing.

Thus, the recommendation of the new statisticians to abandon NHST, understood as the inchoate hybrid commonly employed in psychology, commits the fallacy of the false dichotomy because there exist alternative defensible accounts of NHST (Haig, 2017). The error-statistical perspective is one such attractive alternative.

3.2. Confidence intervals

For the new statisticians, confidence intervals replace p-valued null hypothesis significance testing. Confidence intervals are said to be more informative, and more easily understood, than p values, as well as serving the important scientific goal of estimation, which is preferred to hypothesis testing. Both of these claims are open to challenge. Whether confidence intervals are more informative than statistical hypothesis tests in a way that matters will depend on the research goals being pursued. For example, p values might properly be used to get a useful initial gauge of whether an experimental effect occurs in a particular study, before one runs further studies and reports p values, supplementary confidence intervals, and effect sizes. The claim that confidence intervals are more easily understood than p values is surprising, and is not borne out by the empirical evidence (e.g., Hoekstra et al., 2014). I will speak to the claim about the greater importance of estimation in the next section.

There is a double irony in the fact that the new statisticians criticize NHST for encouraging simplistic dichotomous thinking: As already noted, such thinking is straightforwardly avoided by employing tests of statistical significance properly, whether or not one adopts the error-statistical perspective. For another, the adoption of standard frequentist confidence intervals in place of NHST forces the new statisticians to engage in dichotomous thinking of another kind: Make a decision on whether a parameter estimate is either inside, or outside, its confidence interval.

Error-statisticians have good reason for claiming that their reinterpretation of frequentist confidence intervals is superior to the standard view. The account of confidence intervals adopted by the new statisticians prespecifies a single confidence interval (a strong preference for 0.95 in their case). The single interval estimate corresponding to this level provides the basis for the inference that is drawn about the parameter values, depending on whether they fall inside or outside the interval. A limitation of this way of thinking is that each of the values of a parameter in the interval are taken to have the same evidential, or probative, force – an unsatisfactory state of affairs that results from weak testing. For example, there is no way of answering the relevant questions, 'Are the values in the middle of the interval closer to the true value?', or 'Are they more probable than others in the interval?'

The error-statistician, by contrast, draws inferences about each of the obtained values, according to whether they are warranted, or not, at different severity levels, thus leading to a series of confidence intervals. Mayo (2018) captures the counterfactual logic of severity thinking involved with the following general example: "Were μ less than the 0.995 lower limit, then it is very probable (>0.995) that our procedure would yield a smaller sample mean than 0.6. This probability gives the severity." (p. 195) Clearly, this is a more nuanced and informative assessment of parameter estimates than that offered by the standard view. Details on the error-statistical conception of confidence intervals can be found in Mayo (2018, pp. 189–201), as well as Mayo and Spanos (2011) and Spanos (2014, 2019a, b).

Methodologists and researchers in psychology are now taking confidence intervals seriously. However, in the interests of adopting a sound frequentist conception of such intervals, they would be well advised to replace the new statistics conception of them with their superior error-statistical understanding.

3.3. *Estimation and hypothesis tests*

The new statisticians claim, controversially, that parameter estimation, rather than statistical hypothesis testing, leads to better science – presumably in part because of the deleterious effects of NHST. However, a strong preference for estimation leads Cumming (2012) to aver that the typical questions addressed in science are *what* questions (e.g., "What is the age of the earth?", "What is the most likely sea-level rise by 2012?"). I think that this is a restricted, rather "flattened", view of science where, by implication, explanatory *why* questions and *how* questions (which often ask for information about causal mechanisms) are considered atypical.

Why and how questions are just as important for science as what questions. They are often the sort of questions that science seeks to answer when constructing and evaluating explanatory hypotheses and theories. Interestingly, and at variance with this view, Cumming (Fidler and Cumming, 2014) acknowledges that estimation can be usefully combined with hypothesis testing in science, and that estimation can play a valuable role in theory construction. This is as it should be because science frequently incorporates parameter estimates in precise predictions that are used to assess the hypotheses and theories from which they are derived.

Although it predominantly uses the language of testing, the error-statistical perspective maintains that statistical inference can be employed to deal with both estimation and hypothesis testing problems. It also endorses the view that providing explanations of things is an important part of science and, in fact, advocates piecemeal testing of local hypotheses nested within large-scale explanatory theories.

Despite the generally favorable reception of the new statistics in psychology, it has been subject to criticism by both frequentists (e.g., Sakaluk, 2016), and Bayesians (e.g., Kruschke and Liddell, 2018). However, these criticisms have not occasioned a public response from the principal advocates of the new statistics movement. The error-statistical outlook presents a golden opportunity for those who advocate, or endorse, the new statistics to defend their position in the face of challenging criticism. A sound justification for the promotion and adoption of new statistics practices in psychology requires as much.

4. Bayesian statistics

Despite its early presence, and prominence, in the history of statistics, the Bayesian outlook has taken an age to assert itself in psychology. However, a cadre of methodologists has recently advocated the use of Bayesian statistical methods as a superior alternative to the messy frequentist practice that dominates psychology's research landscape (e.g., Dienes, 2011; Kruschke and Liddell, 2018; Wagenmakers, 2007). These Bayesians criticize NHST, often advocate the use of Bayes factors for hypothesis testing, and rehearse a number of other well-known Bayesian objections to frequentist statistical practice.

Of course, there are challenges for Bayesians from the error-statistical perspective, just as there are for the new statisticians. For example, the frequently made claim that p values exaggerate the evidence against the null hypothesis, but Bayes factors do not, is shown by Mayo not to be the case. She also makes the important point that Bayes factors, as they are currently used, do not have the ability to probe errors and, thus, violate the requirement for severe tests. Bayesians, therefore need to rethink whether Bayes factors can be deployed in some way to provide strong tests of hypotheses through error control.

As with the new statisticians, Bayesians also need to reckon with the coherent hybrid NHST afforded by the error-statistical perspective, and argue against it, rather than the common inchoate hybrids, if they want to justify abandoning NHST. Finally, I note in passing that Bayesians should consider, among other challenges, Mayo's critique of the controversial Likelihood Principle, a principle which ignores the post-data consideration of sampling plans.

4.1. *Contrasts between the Bayesian and error-statistical perspectives*

One of the major achievements of the philosophy of error-statistics is that it provides a comprehensive critical evaluation of the major variants of Bayesian statistical thinking, including the classical subjectivist, "default", pragmatist, and eclectic options within the Bayesian corpus. Whether the adoption of Bayesian methods in psychology will overcome the disorders of current frequentist practice remains to be seen. What is clear from reading the error-statistical literature, however, is that the foundational options for Bayesians are numerous, convoluted, and potentially bewildering. It would be a worthwhile exercise to chart how these foundational options are distributed across the prominent Bayesian statisticians in psychology. For example, the increasing use of Bayes factors for hypothesis testing purposes is accompanied by disorderliness at the foundational level, just as it is in the Bayesian literature more generally. Alongside the fact that some Bayesians are sceptical of the worth of Bayes factors, we find disagreement about the comparative merits of the subjectivist and default Bayesianism outlooks on Bayes factors in psychology (Wagenmakers et al., 2018).

The philosophy of error-statistics contains many challenges for Bayesians to consider. Here, I want to draw attention to three basic features of Bayesian thinking, which are rejected by the error-statistical approach. First, the error-statistical approach rejects the Bayesian insistence on characterizing the evidential relation between hypothesis and evidence in a universal and logical manner in terms of Bayes' theorem. Instead, it formulates the relation in terms of the substantive and specific nature of the hypothesis and the evidence with regards to their origin, modeling, and analysis. This is a consequence of a strong commitment to a piecemeal, contextual approach to testing, using the most appropriate frequentist methods available for the task at hand. This contextual attitude to testing is taken up in Section 5.2, where one finds a discussion of the role different models play in structuring and decomposing inquiry.

Second, the error-statistical philosophy also rejects the classical Bayesian commitment to the subjective nature of prior probabilities, which the agent is free to choose, in favour of the more objective process of establishing error probabilities understood in frequentist terms. It also finds unsatisfactory the turn to the more popular objective, or "default", Bayesian option, in which the agent's appropriate degrees of belief are constrained by relevant empirical evidence. The error-statistician rejects this default option because it fails in its attempts to unify Bayesian and frequentist ways of determining probabilities.

And, third, the error-statistical outlook employs probabilities to measure how effectively *methods* facilitate the detection of error, and how those methods enable us to choose between alternative hypotheses. By contrast, orthodox Bayesians use probabilities to measure *belief* in hypotheses or degrees of confirmation. As noted earlier, most Bayesians are not concerned with error probabilities at all. It is for this reason that error-statisticians will say about Bayesian methods that, without supplementation with error probabilities, they are not capable of providing stringent tests of hypotheses.

4.2. *The Bayesian remove from scientific practice*

Two additional features of the Bayesian focus on beliefs, which have been noted by philosophers of science and statistics, draw attention to their outlook on science. First, Kevin Kelly and Clark Glymour worry that

"Bayesian methods assign numbers to answers instead of producing answers outright." (2004, p. 112) Their concern is that the focus on the scientist's beliefs "screens off" the scientist's direct engagement with the empirical and theoretical activities that are involved in the phenomenology of science. Mayo agrees that we should focus on the scientific phenomena of interest, not the associated epiphenomena of degrees of belief. This preference stems directly from the error-statistician's conviction that probabilities properly quantify the performance of methods, not the scientist's degrees of belief.

Second, Henry Kyburg is puzzled by the Bayesian's desire to "replace the fabric of science … with a vastly more complicated representation in which each statement of science is accompanied by its probability, for each of us." (1992, p.149) Kyburg's puzzlement prompts the question, 'Why should we be interested in each other's probabilities?' This is a question raised by David Cox about prior probabilities, and noted by Mayo (2018).

This Bayesian remove from science contrasts with the willingness of the error-statistical perspective to engage more directly with science. Mayo is a philosopher of science as well as statistics, and has a keen eye for scientific practice. Given that contemporary philosophers of science tend to take scientific practice seriously, it comes as no surprise that she brings it to the fore when dealing with statistical concepts and issues. Indeed, her error-statistical philosophy should be seen as a significant contribution to the so-called *new experimentalism*, with its strong focus, not just on experimental practice in science, but also on the role of statistics in such practice. Her discussion of the place of frequentist statistics in the discovery of the Higgs boson in particle physics is an instructive case in point.

Taken together, these just-mentioned points of difference between the Bayesian and error-statistical philosophies constitute a major challenge to Bayesian thinking that methodologists, statisticians, and researchers in psychology need to confront.

4.3. Bayesian statistics with error-statistical foundations

One important modern variant of Bayesian thinking, which now receives attention within the error-statistical framework, is the *falsificationist Bayesianism* of Andrew Gelman, which received its major formulation in Gelman and Shalizi (2013). Interestingly, Gelman regards his Bayesian philosophy as essentially error-statistical in nature – an intriguing claim, given the anti-Bayesian preferences of both Mayo and Gelman's co-author, Cosma Shalizi. Gelman's philosophy of Bayesian statistics is also significantly influenced by Popper's view that scientific propositions are to be submitted to repeated criticism in the form of strong empirical tests. For Gelman, best Bayesian statistical practice involves formulating models using Bayesian statistical methods, and then checking them through hypothetico-deductive attempts to falsify and modify those models.

Both the error-statistical and neo-Popperian Bayesian philosophies of statistics extend and modify Popper's conception of the hypothetico-deductive method, while at the same time offering alternatives to received views of statistical inference. The error-statistical philosophy injects into the hypothetico-deductive method an account of statistical induction that employs a panoply of frequentist statistical methods to detect and control for errors. For its part, Gelman's Bayesian alternative involves formulating models using Bayesian statistical methods, and then checking them through attempts to falsify and modify those models. This clearly differs from the received philosophy of Bayesian statistical modeling, which is regarded as a formal inductive process.

From the wide-ranging error-statistical evaluation of the major varieties of Bayesian statistical thought on offer, Mayo concludes that Bayesian statistics needs new foundations: In short, those provided by her error-statistical perspective. Gelman acknowledges that his falsificationist Bayesian philosophy is underdeveloped, so it will be interesting to learn how its further development relates to Mayo's error-statistical perspective. It will also be interesting to see if Bayesian thinkers in psychology engage with Gelman's brand of Bayesian thinking. Despite the appearance of his work in a prominent psychology journal, they have yet to do so. However, Borsboom and Haig (2013) and Haig (2018) provide sympathetic critical evaluations of Gelman's philosophy of statistics.

It is notable that in her treatment of Gelman's philosophy, Mayo emphasizes that she is willing to allow a decoupling of statistical outlooks and their traditional philosophical foundations in favour of different foundations, which are judged more appropriate. It is an important achievement of Mayo's work that she has been able to consider the current statistics wars without taking a particular side in the debates. She achieves this by examining methods, both Bayesian and frequentist, in terms of whether they violate her minimal severity requirement of "bad evidence, no test".

5. The error-statistical perspective and the nature of science

As noted at the outset, the error-statistical perspective has made significant contributions to our philosophical understanding of the nature of science. These are achieved, in good part, by employing insights about the nature and place of statistical inference in experimental science. The achievements include deliberations on important philosophical topics, such as the demarcation of science from non-science, the underdetermination of theories by evidence, the nature of scientific progress, and the perplexities of inductive inference. In this article, I restrict my attention to two such topics: The process of falsification and the structure of modeling.

5.1. Falsificationism

The best known account of scientific method is the so-called hypothetico-deductive method. According to its most popular description, the scientist takes an existing hypothesis or theory and tests indirectly by deriving one or more observational predictions that are subjected to direct empirical test. Successful predictions are taken to provide inductive confirmation of the theory; failed predictions are said to provide disconfirming evidence for the theory. In psychology, NHST is often embedded within such a hypothetico-deductive structure and contributes to weak tests of theories.

Also well known is Karl Popper's falsificationist construal of the hypothetico-deductive method, which is understood as a general strategy of conjecture and refutation. Although it has been roundly criticised by philosophers of science, it is frequently cited with approval by scientists, including psychologists, even though they do not, indeed could not, employ it in testing their theories. The major reason for this is that Popper does not provide them with sufficient methodological resources to do so.

One of the most important features of the error-statistical philosophy is its presentation of a falsificationist view of scientific inquiry, with error statistics serving an indispensable role in testing. From a sympathetic, but critical, reading of Popper, Mayo endorses his strategy of developing scientific knowledge by identifying and correcting errors through strong tests of scientific claims. Making good on Popper's lack of knowledge of statistics, Mayo shows how one can properly employ a range of, often familiar, error-statistical methods to implement her all-important severity requirement. Stated minimally, and informally, this requirement says, "A claim is severely tested to the extent that it has been subjected to and passes a test that probably would have found flaws, were they present." (Mayo, 2018, p. xii) Further, in marked contrast with Popper, who deemed deductive inference to be the only legitimate form of inference, Mayo's conception of falsification stresses the importance of inductive, or content-increasing, inference in science. We have here, then, a viable account of falsification, which goes well beyond Popper's account with its lack of operational detail about how to construct strong tests. It is worth noting that the error-statistical stance offers a constructive interpretation of Fisher's oft-cited remark that the null hypothesis is never proved, only possibly disproved.

5.2. A hierarchy of models

In the past, philosophers of science tended to characterize scientific inquiry by focusing on the general relationship between evidence and theory. Similarly, scientists, even today, commonly speak in general terms of the relationship between data and theory. However, due in good part to the labors of experimentally-oriented philosophers of science, we now know that this coarse-grained depiction is a poor portrayal of science. The error-statistical perspective is one such philosophy that offers a more fine-grained parsing of the scientific process.

Building on Patrick Suppes' (1962) important insight that science employs a hierarchy of models that ranges from experimental experience to theory, Mayo's (1996) error-statistical philosophy initially adopted a framework in which three different types of models are interconnected and serve to structure error-statistical inquiry: Primary models, experimental models, and data models. Primary models, which are at the top of the hierarchy, break down a research problem, or question, into a set of local hypotheses that can be investigated using reliable methods. Experimental models take the mid-positon on the hierarchy and structure the particular models at hand. They serve to link primary models to data models. And, data models, which are at the bottom of the hierarchy, generate and model raw data, put them in canonical form, and check whether the data satisfy the assumptions of the experimental models. It should be mentioned that the error-statistical approach has been extended to primary models and theories of a more global nature (Mayo and Spanos, 2010) and, now, also includes a consideration of experimental design and the analysis and generation of data (Mayo, 2018).

This hierarchy of models facilitates the achievement of a number of goals that are important to the error-statistician. These include piecemeal strong testing of local hypotheses rather than broad theories, and employing the model hierarchy as a structuring device to knowingly move back and forth between statistical and scientific hypotheses. The error-statistical perspective insists on maintaining a clear distinction between statistical and scientific hypotheses, pointing out that psychologists often mistakenly take tests of significance to have direct implications for substantive hypotheses and theories.

6. The philosophy of statistics

A heartening attitude that comes through in the error-statistical corpus is the firm belief that the philosophy of statistics is an important part of statistical thinking. This emphasis on the conceptual foundations of the subject contrasts markedly with much of statistical theory, and most of statistical practice. It is encouraging, therefore, that Mayo's philosophical work has influenced a number of prominent statisticians, who have contributed to the foundations of their discipline. Gelman's error-statistical philosophy canvassed earlier is a prominent case in point. Through both precept and practice, Mayo's work makes clear that philosophy can have a direct impact on statistical practice. Given that statisticians operate with an implicit philosophy, whether they know it or not, it is better that they avail themselves of an explicitly thought-out philosophy that serves their thinking and practice in useful ways. More particularly, statistical reformers recommend methods and strategies that have underlying philosophical commitments. It is important that they are identified, described, and evaluated.

The tools used by the philosopher of statistics in order to improve our understanding and use of statistical methods are considerable (Mayo, 2011). They include clarifying disputed concepts, evaluating arguments employed in statistical debates, including the core commitments of rival schools of thought, and probing the deep structure of statistical methods themselves. In doing this work, the philosopher of statistics, as philosopher, ascends to a meta-level to get purchase on their objects of study. This second-order inquiry is a proper part of scientific methodology.

It is important to appreciate that the error-statistical outlook is a scientific methodology in the proper sense of the term. Briefly stated, methodology is the interdisciplinary field that draws from disciplines that include statistics, philosophy of science, history of science, as well as indigenous contributions from the various substantive disciplines. As such, it is the key to a proper understanding of statistical and scientific methods. Mayo's focus on the role of error statistics in science is deeply informed about the philosophy, history, and theory of statistics, as well as statistical practice. It is for this reason that the error-statistical perspective is strategically positioned to help the reader to go beyond the statistics wars.

7. Conclusion

The error-statistical outlook provides researchers, methodologists, and statisticians with a distinctive and illuminating perspective on statistical inference. Its Popper-inspired emphasis on strong tests is a welcome antidote to the widespread practice of weak statistical hypothesis testing that still pervades psychological research. More generally, the error-statistical standpoint affords psychologists an informative perspective on the nature of good statistical practice in science that will help them understand and transcend the statistics wars into which they have been drawn. Importantly, psychologists should know about the error-statistical perspective as a genuine alternative to the new statistics and Bayesian statistics. The new statisticians, Bayesians statisticians, and those with other preferences should address the challenges to their outlooks on statistics that the error-statistical viewpoint provides. Taking these challenges seriously would enrich psychology's methodological landscape.

Declaration of Competing Interest

The author has no conflict of interest to declare.

References

Borsboom, D., Haig, B.D., 2013. How to practice Bayesian statistics outside the Bayesian church: what philosophy for Bayesian statistical modelling? Br. J. Math. Stat. Psychol. 66, 39–44.
Calin-Jageman, R.J., Cumming, G., 2019. The new statistics for better science: ask how much, how uncertain, and what else is known. Am. Statistician 73 (No. S1), 271–280.
Cumming, G., 2012. Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. Routledge.
Cumming, G., 2014. The new statistics: why and how. Psychol. Sci. 25, 7–29.
Dienes, Z., 2011. Bayesian versus orthodox statistics: which side are you on? Perspect. Psychol. Sci. 6, 274–290.
Eich, E., 2014. Business not as usual. Psychol. Sci. 25, 3–6.
Fidler, F., Cumming, G., 2014. Yes, but don't underestimate estimation: reply to Morey, rouder, verhagen, and Wagenmakers (2014). Psychol. Sci. 25, 1291–1292.
Gelman, A., Shalizi, C.R., 2013. Philosophy and the practice of Bayesian statistics. Br. J. Math. Stat. Psychol. 66, 8–38.
Haig, B.D., 2017. Tests of statistical significance made sound. Educ. Psychol. Meas. 77, 489–506.
Haig, B.D., 2018. The Philosophy of Quantitative Methods. Oxford University Press.
Hoekstra, R., Morey, R.D., Rouder, J.N., Wagenmakers, E.-J., 2014. Robust misinterpretation of confidence intervals. Psychonomic Bull. Rev. 21, 1157–1164.
Kelly, K.T., Glymour, C., 2004. Why probability does not capture the logic of scientific justification. In: Hitchcock, C. (Ed.), Contemporary Debates in Philosophy of Science. Blackwell, pp. 94–114.
Kruschke, J.K., Liddell, T.M., 2018. The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bull. Rev. 25, 178–206.
Kyburg, H., 1992. The scope of Bayesian reasoning. In: PSA:Proceedings of the Biennial Meeting of the Philosophy of Science Association, vol. 2. University of Chicago Press, pp. 139–152.
Mayo, D.G., 1996. Error and the Growth of Experimental Knowledge. University of Chicago Press.
Mayo, D.G., 2011. Statistical science and the philosophy of science: where do/should they meet in 2011 (and beyond)? Rationality, Markets, and Morals 2, 79–102.
Mayo, D.G., 2018. Statistical Inference as Severe Testing: How To Get beyond the Statistics Wars. Cambridge University Press.

Mayo, D.G., Spanos, A. (Eds.), 2010. Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science. Cambridge University Press.
Mayo, D.G., Spanos, A., 2011. Error statistics. In: Bandyopadhyay, P.S., Forster, M.R. (Eds.), Handbook of Philosophy of Science, Philosophy of Statistics, vol. 7. Elsevier, pp. 153–198.
Sakaluk, J., 2016. Exploring small, confirming big: an alternative system to the new statistics for advancing cumulative and replicable psychological research. J. Exp. Soc. Psychol. 66, 47–54.
Spanos, A., 2014. Recurring controversies about *P* values and confidence intervals revisited. Ecology 95, 645–651.
Spanos, A., 2019a. Probability Theory and Statistical Inference: Empirical Modeling with Observational Data. Cambridge University Press.
Spanos, A., 2019b. What foundations for statistical modeling and inference? Economia 9, 33–860.
Suppes, P., 1962. Models of data. In: Nagel, E., Suppes, P., Tarski, A. (Eds.), Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress. Stanford University Press, pp. 252–261.
Wagenmakers, E.-J., 2007. A practical solution to the pervasive problems of *p* values. Psychonomic Bull. Rev. 14, 779–804.
Wagenmakers, E.-J., et al., 2018. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. Psychon. Bull. Rev. 25, 35–57.