

DEBORAH G. MAYO & ARIS SPANOS, eds. 2009. *Error and Inference*. Cambridge: Cambridge University Press.

*Error and Inference* focuses on the error-statistical philosophy of science (ESP) put forward by Deborah Mayo and Aris Spanos (MS). Chapters 1, 6 and 7 are mainly written by MS (partly with the statistician David Cox), whereas Chapters 2-5, 8, and 9 are driven by the contributions of other authors. There are responses to all these contributions at the end of the chapters, usually written by Mayo.

The structure of the book with the responses at the end of each chapter is a striking feature. The critical contributions enable a very lively discussion of ESP. On the other hand always having the last word puts Mayo and Spanos in a quite advantageous position. Some of the contributors may have underestimated Mayo's ability to make the most of this advantage.

Central to ESP are the issues of probing scientific theories objectively by data, and Mayo's concept of "severe testing" (ST). ST is based on a frequentist interpretation of probability, on conventional hypothesis testing and the associated error probabilities. ESP advertises a "piecemeal" approach to testing a scientific theory, in which various different aspects, which can be used to make predictions about data, are subjected to hypothesis tests. A statistical problem with such an approach is that failure of rejection of a null hypothesis  $H_0$  does not necessarily constitute evidence in favour of  $H_0$ . The space of probability models is so rich that it is impossible to rule out all other probability models.

This motivates "severity". In case that  $H_0$  is not rejected by a test  $T$  with observed data  $x$ , severity is defined as the probability that  $T$  would have produced a result according less well with  $H_0$  than  $x$  if  $H_0$  were false.

If this probability is high, it is said that  $H_0$  passed a severe test. In a frequentist setup, as opposed to a Bayesian one, probabilities are not assigned to epistemic statements like "the theory/hypothesis is true", but only to observable events assuming a true underlying probability model. If a theory is supported by strong evidence according to ESP, this does therefore not imply that the theory has a high probability of being true.

" $H_0$  is false" above refers to assuming the "true" underlying distribution in some distance of  $H_0$ . "Distance" is defined in terms of the alternative model against which  $H_0$  is tested. Distributions not belonging to this model cannot be ruled out. The essence of the "piecemeal" is that a scientific theory gives rise to various different null hypotheses, and various tests of every single one. Additionally to testing  $H_0$  within a certain model, the model assumptions can be tested ("misspecification testing").

Chapter 1 by Mayo and the introductory section by MS give an overview of ESP. A major theme is how a piecemeal of local and statistically testable hypotheses can be related to a large-scale theory. Bayesian approaches are criticised for the difficulty to model a "catch-all alternative" of a theory under investigation.



In Chapter 2 and 3, Alan Chalmers and Alan Musgrave argue from two different points of view (emphasising the role of theories in science, and critical rationalism) that ST is too strong a demand for high-level theories. Mayo, in her responses, wonders whether Chalmers' arguments reflect too much of a "desire to have things settled" instead of accepting that a theory is reliable only as far as it has been severely tested, and she denies the necessity of "believing" large-scale theories in order to make use of them. She emphasises that ESP is agnostic about scientific realism.

John Worrall, in Chapter 4, states that data needs to be "use-novel" in order to constitute evidence in favor of a hypothesis. ESP sometimes requires a double-use of data, e.g., when the same data is used for model misspecification testing and for inference within a model. Mayo distinguishes situations where use-novelty is required (to prevent hypotheses to be constructed in such a way that the data have no chance to contradict them) from where it is not required (where there is no danger of confirming wrong hypotheses).

In Chapter 5, Peter Achinstein defends the assignment of epistemic probabilities to hypotheses. Mayo responds that Achinstein's examples rely on a wrong interpretation of error probabilities. Chapter 6 by Spanos is devoted to theory testing in economics from an ESP perspective. He discusses the interplay between theory, data and statistical modeling, and gives a historical overview of the too often neglected role of empirical data in economical theory testing.

Section I of Chapter 7 (as Section II written by Cox and Mayo) discussed issues in the philosophy of statistics. Section II starts off with a paragraph on objectivity ("independent of our beliefs, biases and interests"). Sufficiency and conditioning in frequentist modelling are discussed and it is denied that there can be really objective Bayesian prior distributions. Section III by Mayo contains a remarkable argument against a notorious statement of Birnbaum (*JASA* 57 [1962]) that the strong likelihood principle (SLP) is entailed by two principles that frequentists share. It has been claimed that this means that frequentists should not use p-values and Neyman-Pearson testing, both of which violate the SLP. Mayo shows that Birnbaum's argument has two premises that cannot both be true at the same time. Spanos argues in Section IV that it is possible to test hypotheses in a logical order in order to arrive at a confirmed model. He also states a condition which makes it possible to "double-use" data for misspecification testing. However, this condition is often not fulfilled, and Spanos does not mention that fulfillment of the condition itself cannot be tested in this way.

Chapter 8 by Clark Glymour on connections between causal explanations and ST is broadly in agreement with MS. Mayo emphasises that she promotes "twin goals of severity and informativeness". Spanos connects the themes of misspecification testing and structural vs. statistical models to causality. In Chapter 9, Larry Laudan writes about an inconsistency in the law system regarding the burden of proof in case of an "affirmative defence". At first sight this seems to be remote from the general topic of the book, but surprisingly Laudan comes up with the only objection against ESP in the book that Mayo does not address convincingly. He points out that clear acceptance rules and standards of proof are lacking in the philosophy of science. Par-

ticularly it is not clear what is required to test a theory “severely enough”. In fact, subjective decisions about cutoff values for error probabilities and distances of alternatives from the  $H_0$  are required, and one can wonder whether calling a theory “severely tested” is as objective as MS imply. Mayo’s response concerns the dependence of such decisions on a “cost of errors”, which she calls a “policy or management issue”.

A major quality of the book not shared by many philosophical accounts of such issues is the statistical competence of MS backed up by the rich data analytic experience of Cox and Spanos. Mayo’s logical elaborations are sharp and convincing. In various places she rises metaphilosophical issues beyond standard ESP issues, e.g., the role of counterexamples in the philosophy of science. Spanos is well readable, though somewhat repetitive at times.

My personal concern with ESP is that its proponents seem to be overoptimistic about what it can achieve. I find myself in broad agreement with MS regarding the practical statistical implications, but less so with their philosophical interpretation. MS are somewhat ambiguous regarding the “truth” of models. On one hand, they are obviously aware that all probability models are idealisations. On the other hand, they often argue as if there is a true model, which can be more or less reliably approximated using the proposed modelling/misspecification routines.

MS suggest that models should be tested severely against model misspecification, but there are no examples in which severity calculations are actually carried out for model misspecification tests. Such calculations may often not be possible, and parametric inference may still fail if misspecification tests pass the model. To be fair, MS are aware of these issues.

Compared with some of the other contributors and other philosophers of science I have read on similar issues, particularly Mayo is laudably sceptical and modest about what theory testing can achieve. Still I believe that the merits of ESP could be better explained in pragmatist terms than in terms of “objective truth”. The Bayesian perspective is heavily criticised in the book and it is a pity that there is no contribution by a modern Bayesian statistician. Overall, this is a very valuable and thought-provoking book which makes a strong case in favour of ESP.

Christian Hennig  
University College London  
chrish@stats.ucl.ac.uk

JEROEN VAN BOUWEL, ed. 2009. *The Social Sciences and Democracy*. London: Palgrave Macmillan.

The idea behind *The Social Sciences and Democracy* is an exciting one: Pulling together scholars from philosophy of science, the social studies of science and historians of thought to explore the myriad of ways in which democracy and the social sciences are related. The connections drawn range from the analysis of groups of scientists as democratic communities, to the question of what role social scientists should play in a democratic society, on to how social scientific research can inform our normative