Pergamon

Stud. Hist. Phil. Sci., Vol. 29, No. 3, pp. 465–479, 1998

Printed in Great Britain
0039–3681/98 $19.00+0.00

Essay Review

Exploiting Errors

*Giora Hon**

Mayo, Deborah G., *Error and the Growth of Experimental Knowledge* (Chicago and London: Chicago University Press, 1996), xvi + 494 pp., ISBN 0-226-51197-9 Hardback $74.00; 0-226-51198-7 Paperback $29.95.

> Any planned inquiry in which there is a deliberate and reliable argument from error may be said to be experimental. D. G. Mayo (1996, p. 7).

The title of Deborah Mayo's book, *Error and the Growth of Experimental Knowledge*, takes its cue from the subtitle of Karl Popper's celebrated book, *Conjectures and Refutations: the Growth of Scientific Knowledge* (1963), which, according to Popper, revolves around one simple theme: 'the thesis that *we can learn from our mistakes*' (Popper, 1974, vii; original emphasis). Mayo's reference to experimental, rather than scientific, knowledge prompts the reader to think further of Thomas Kuhn's notion of normal science—the day-to-day business of science. That in turn triggers a thought of the way the conducting of experiments informs the practice of science as it has been recently analysed by New Experimentalists like Ian Hacking and Peter Galison. Such are one's associations.

It therefore comes as a surprise that a book which bears a title containing the philosophically pregnant words 'error and the growth of knowledge' and is published in a series dedicated to 'Science and Its Conceptual Foundations' (under the editorship of David L. Hull), should belong not in a philosophical milieu but rather in a scientific, technical one. The librarian of the Library of Congress catalogued Mayo's book as QA275.M347 1996. In my university library the book is found flanked on its left by Longley's *Least Squares Computations Using Orthogonalization Methods* (1984) and Biemer *et al.*'s *Measurement Errors in Surveys* (1991). On its right the book is flanked by Shchigolev's *Mathematical Analysis of Obser-*

* Department of Philosophy, Haifa University, Mount Carmel, Haifa 31905, Israel

PII: S0039-3681(98)00006-5

vations (1965) and Topping's classical textbook, *Errors of Observation and Their Treatment* (1969). Is this the appropriate company for the book under review? Is QA275 its natural habitat? Or did the librarian make a mistake? Why isn't the book in the company of books written by philosophers of science like Popper and Kuhn or by New Experimentalists like Hacking and Galison? Their books are placed a few shelves away, in the BD, Q, and QC sections. QA275 is rather the habitat of books concerned with what is technically termed experimental design, and the mathematical subject of error analysis.

To make the situation more confusing, the librarian recorded in the publication data that the book addresses the following subjects: '1. Error analysis (Mathematics) 2. Bayesian statistical decision theory. 3. Science—Philosophy' (title-page, reverse). What is then the subject of this book? Is the book a technical manual for researchers, one of many guides to experimental design and error analysis of the QA275 type, or did the author have in mind an audience of philosophers of science? The blurb on the back cover is not helpful: 'The book will be important for philosophers and methodologists of science as well as for researchers'. Was an error committed in classifying Mayo's book?

Error is a multifarious epistemological phenomenon which consists essentially in a separation between elements that may be either concrete or abstract. In the former case material objects that have been intended to coincide are in fact at a distance; in the latter, an incoherence produces a breach between propositions that are assumed to concur one with another.[1] The Latin *erro* means 'A. In general, to wander or stray about, to wander up and down, to rove... B. In particular, to miss the right way, to lose one's self, go astray...' (Lewis and Short, 1966, p. 657; cf. Bates, 1996, p. 312). Error is an expression of divergence whose mark is discrepancy—a discrepancy which emerges from a procedure of evaluation against a chosen standard. The nature of this discrepancy, the reason for its occurrence, how to treat it and what can be learnt from it once it has been perceived and comprehended, constitute the vast subject of the problem of error.

Missing the target is the most concrete and, historically, ancient of all types of error. Indeed, the Greek term for error, *hamartia*, is regarded as a 'falling short' or a 'missing the mark', that is, a failure; its opposite is *tynchano*—to hit the mark. The battlefield in the *Iliad* provides an admirable background for numerous cases of hitting and missing (Bremer, 1969, p. 30).

> Then strong Diomedes answered, not frightened before him: 'You did not hit me, you missed [erred, G.H.], but I do not think that you two will go free until one or the other of you has fallen...' (Lattimore, 1965, V. 286–288).

Clearly, a wide-shot—the fundamental error—can be a matter of life and death. This elementary fact has served as a metaphor in the epistemological realm. Searching for the definition of knowledge, Socrates discussed with Theaetetus the conditions of error and observed that,

[1]Notice that lies too consist of separation but there the gap is intentional.

> Like a bad archer, you may shoot to one side and miss the mark—which is indeed another phrase we use for error. (Cornford, 1957, p. 125; *Theaetetus*, 194).

Hamartia, the term for missing the target, evolved further in the *Poetics* of Aristotle into the notion of 'tragic flaw', or 'tragic error'. It also became the expression for missing the golden mean, and *tychano* for hitting the mean, an idea used as a metaphor by the Stoa to refer to lacking or possessing the balance of moral rectitude (Bremer, 1969, pp. 52–56). Thus, missing, erring and failing in moral rectitude became entangled right at the inception of Western thought. Error has since then been associated with both 'knowing that', i.e. propositional knowledge, and 'knowing how', i.e. tacit, skilful knowledge. Surprisingly, this rich and important problem has not received a systematic historical and philosophical account. The problem is in a state of disgraceful neglect.

'The problem of error is one of philosophy's very serious and crucial problems', observed Koyré commenting on Plato's discussion of the nature of error in the *Theaetetus* (Koyré, 1945, p. 40, note 9). Koyré however made this observation in an inconsequential footnote and did not elaborate. Consider, for another illustration of this neglect, Quine's recent book, *From Stimulus to Science*. Quine opens his book with this very problem.

> We and other animals notice what goes on around us. This helps us by suggesting what we might expect and even how to prevent it, and thus fosters survival. However, the expedient works only imperfectly. There are surprises, and they are unsettling. How can we tell when we are right? We are faced with the problem of error. These are worries about our knowledge of the external world. To deal with them we have had to run inward and seek knowledge *of* knowledge (Quine, 1995, p. 1).

The reader of Quine's book soon realizes, however, that the problem of error is being used here as a literary device rather than as a real philosophical issue. The problem receives no explicit attention in the remaining parts of the book.

Notwithstanding, it appears that historians and philosophers of science are increasingly paying attention to the vast and varied problem of error.[2] This growing concern with the concept of error is connected to the shift of attention which the history and the philosophy of science are currently undergoing. The actual nitty-gritty practice of scientific research and the social foundation of the resulting knowledge have become legitimate objects of historical and philosophical studies as much as the conceptual content itself (see, e.g., Galison, 1987; Shapin, 1994). Gone are the days when one could opine casually that 'once [errors of measurement and other forms of experimental error]... have been discounted, our attention can turn to the logico-mathematical structure' (Sellars, 1961, p. 73). The occurrence of errors, especially in experimentation, constitutes a permanent feature which deserves proper attention. It has been realized by now that the problem of error is

[2]See, e.g., Hon (1989a); the contrast between Hon (1987a) and Buchwald (1995a, b); Krüger *et al.* (1989), Section III: Uncertainty; Mirowski (1994, 1995); Stigler (1986); the numerous publications of Sheynin; for a discussion in the cultural-political realm, Bates (1996); for a rich annotated bibliography of the problem of error, Swijtink (forthcoming).

not incidental to the pursuits of science. As Mellor opined, error cannot be treated as 'a tiresome but trivial excrescence on the neat deductive structure of science' (Mellor, 1967, p. 6).

Is the book under review an exemplar of the new outlook in philosophy of science? *Prima facie* this is the case. The author is vigorously critical of the orthodoxies of Popper and Kuhn and vehemently rejects the popular Bayesian approach in philosophy of science. Building on a classical statistical tradition (e.g., pp. x, 10, 337; see however Mirowski, 1995, p. 547, note 10), the author develops a sophisticated technical machinery with which she underpins a non-Bayesian philosophy of science. Mayo sides with Peirce, Neyman and Pearson; she improves upon their ideas and arguments against their common opponents—the Bayesians. She christens her position 'error-statistical philosophy of experiment' (e.g. pp. 410, 442, 457, 464), because the chief feature that her approach retains from the Neyman–Pearson statistical methods is the centrality of error probabilities (pp. x–xi). The demand that it is necessary to take into account the error probabilities of experimental procedure in order to determine what inferences are licensed by data, is the principal element that fundamentally distinguishes, according to the author, her approach from others (p. 442).

The statistical methods which Mayo offers are designed not only to stabilize experimental knowledge and explain its growth but also to grapple with other issues pertinent to philosophy of science, such as the Duhem problem (pp. 103, 106–109). Indeed, in referring to an error-statistical philosophy of science, the author has in mind the various ways in which statistical methods based on error probabilities may be used in the philosophy of science generally. Mayo wishes to impress upon the reader that the error-statistical philosophy of science which she has developed has a structure and a logic, so that its parts hang together to provide a full-bodied philosophy. In her view, this philosophy presents a viable alternative to the Bayesian Way (pp. 442–444) and may provide

> the general framework and tools for carrying out the main goals of the New Experimentalist program (p. 456).

On a most optimistic note, Mayo concludes triumphantly that,

> [i]f there is experimental knowledge to be had of a phenomenon, then it will be detectable by means of... [the proposed] methods. The ability to make successful inductions, our success in obtaining experimental knowledge, is explained by the error-statistical properties of our methods. We make progress in experimental knowledge—experimental knowledge grows—because we have methods that are manifestly adequate for learning from errors (p. 464).

On this account, errors and the statistical methods for treating them have become the tools for building the body of knowledge we call science. Has error then gained in this rejuvenated classical statistical approach the respect due to it?

Error covers multiple sins. It is a multifarious epistemological phenomenon of great breadth and depth. To be sure, error-statistical analysis is a powerful tool much needed in the technical realm of the reduction of data and it can undoubtedly throw light on methodological issues, but it cannot do philosophical justice, at least not on its own, to such complex concepts as error and experiment. Mayo's book is not about error but about error probabilities, and the notion of experimental knowledge it develops is rather the knowledge of the probabilities of specified outcomes in some series of experiments (p. 12).

The central problem which Mayo addresses is how to link experimental data to primary theoretical hypotheses. It is commonly known that data gathered from experiment are corrupted by various kinds of error introduced by intermediary processes of observation and measurement. Moreover, data are finite and discrete, while primary hypotheses may refer to an infinite number of cases and involve continuous quantities such as weight and temperature (p. 132). Nevertheless, since the mandatory linkage between data and hypothesis is the only game in town which deserves the appellation scientific, one must use experimental data to assess the values of theoretical quantities. How is this done? How are we to proceed? This is an age-old problem which Mayo addresses successfully with much vigour and stamina. She intends to use ideas from statistics to obtain a philosophical understanding of reasoning in science (p. 337).

Mayo follows a well-known distinction (see, e.g., Hacking, 1965, Chapter VI). She distinguishes between 'evidential-relationship' (E-R) approaches and 'testing' approaches. E-R approaches emerged naturally from what was traditionally thought to be required by a 'logic' of confirmation or induction. They commonly seek quantitative measures of the bearing of evidence on hypotheses. Specifically, these quantities are probabilities or other measures of support or credibility assigned to hypotheses. By contrast, the testing approaches focus, according to Mayo, on finding general methods or procedures of testing with certain good properties. The quantities in testing approaches refer only to properties of methods, for example, of testing or of estimation procedures (p. 72). The Bayesian inference is a typical E-R approach, while testing approaches include non-Bayesian approaches such as Popperian corroboration, Fisherian statistics and Neyman–Pearson statistics as well as entirely qualitative non-Bayesian approaches (p. 73).

Mayo notes the popularity of E-R approaches amongst philosophers of science. She explains that the great appeal of the E-R is due to the confidence these approaches give and that they are, as it were, clean: the E-R view is modeled on deductive logic, only with probabilities. The task here is to determine, for given evidence and hypotheses, how well the evidence confirms or supports hypotheses. By contrast, in the testing approach one has to eschew grand and unified schemes and opt for a hodgepodge of methods that are truly ampliative. Here statistical tools are applied as protection from the many ways in which one can be misled by data as well as by one's own beliefs and desires. The tools are used for developing strategies, for collecting data and for efficiently checking an assortment of errors. Clearly, the information that is of interest to the E-R proponent is very different from that which is relevant to the advocate of the testing approach (pp. 337–338).

Specifically, in the Bayesian approach, what one learns about a hypothesis H

from evidence *e* is measured by the conditional probability of *H* given *e* using Bayes's theorem. The cornerstone of the Bayesian approach is the use of prior probability assignments to hypotheses, generally interpreted as an agent's subjective degrees of belief. The prior probability assignments might be individually idiosyncratic, but they ultimately obey the Kolmogorov axioms (Mirowski, 1994, p. 566). By contrast, the methods and models of classical and Neyman–Pearson statistics (e.g. statistical significance tests, confidence interval methods) reject the use of prior probabilities as these could not be based on objective frequencies. Instead, a frequentist theory of probability is presupposed; it enters by way of characterizing the experimental or testing process itself—expressing how reliably the process discriminates between alternative hypotheses (p. x; cf. Mirowski, *ibid.*).

Mayo resolutely advocates the testing approach. Indeed, the book revolves around the notion of a severe testing process. Mayo substantiates and persuasively defends an approach based on severe testing, which underwrites the error-statistical philosophy of science she proposes. Hence the need to be very clear on the fundamental differences between Bayesian and error-probability approaches (p. 337), a demand which Mayo fulfills with flying colours. Her primary aim is to show how a number of disputes in philosophy of science reflect the difference between the approaches based on the evidence-relation and testing, and to build an account of experimental learning based on the error-statistical approach. Mayo is concerned to show that the error approach is at the heart of the widespread applications of statistical ideas in scientific inquiry, and that it offers a fruitful basis for a philosophy of experiment (p. 361). In other words, the author seeks to exhibit realistically the way scientific inquiries are actually conducted; she is not interested in coming up with some vacuous precepts divorced from the practice of science (p. 362). Thus,

> to restore the role of empirical data as an objective constraint and adjudicator in science, we need to study the actual experimental processes and reasoning that are used to arrive at data (p. 60).

Indeed, Mayo perceptively points out that the study of the relation between evidence and hypotheses solely in terms of logical relationships ignores completely all the deliberate and active intervention in which the experimenter is engaged (p. 212).

> In focusing too exclusively on the appraisal of global theories, philosophers have overlooked how positive grounds are provided for local hypotheses, namely, whenever evidence counts as having severely tested them. By attempting to talk about data and hypotheses in some general way, apart from the specific context in which the data and hypothesis are generated, modeled, and analyzed to answer specific questions, philosophers have missed the power of such a piecemeal strategy, and underdetermination arguments have flourished (p. 213).

For example, what matters to the Bayesian is not so much the experiment itself but rather what is the likelihood function after experimentation. By contrast, Mayo convincingly demonstrates that Perrin, the celebrated French experimenter, certainly did not attempt in his studies of Brownian motion to state prior probabilities and multiply them by likelihoods to yield posteriors (p. 232; cf. Chapter 7). In sum, Mayo advises us not to follow the Bayesian Way, but rather the path of the classical statistician and to search, as Pearson put it,

> *for a way of expressing in mathematical terms what appeared... to be the requirements of the scientist in applying statistical tests to his data* (p. 381; Mayo's emphasis).

The application of statistical tests is the key idea, and as Pearson reported,

> from the start we [i.e. Neyman and Pearson] shared Professor Fisher's view that in scientific enquiry, a statistical test is 'a means of learning' (p. 382).

On this account one learns in science not from how much the evidence confirms the hypothesis tested, but rather from *how discordant* evidence shows a given model to be in a specified respect. Learning from experiments requires not some update of the probability assignment that one starts out with, but deliberate and often devious methods of testing with which one builds, corrects, and fills out a model (p. 212, 433).

The theme of learning from error thus plays a central role in the experimental programme which Mayo develops. She demonstrates how the famous Popperian thesis of 'conjecture and refutation'—Popper's way of learning from mistakes—does not stand up to criticism. According to Mayo, Popper's account falls far short of showing how reliable knowledge is obtained from experiment or how that knowledge grows. In spite of the fact that both are in the testing-approach camp, Mayo's account does not find a home in the Popperian framework. It is quite at home, however, within the experimental framework of another philosopher who also developed an account wherein scientific inference is based on learning from error and error correction, namely, C. S. Peirce (p. 412). Mayo finds in the Peircean error-correcting justification of induction, the very justification she needs for her error-statistical methods of science.

> The justification for these methods lies in their ability to control error probabilities, hence sustain learning from error, hence provide for the growth of experimental knowledge (p. 413).

To understand the nature and growth of experimental knowledge, one must look, according to the author, to normal, standard testing which is in her view nothing else than the Kuhnian conception of normal science.

Mayo, however, is as critical of Kuhn as of Popper. She does not accept Kuhn's supposition that there are two kinds of scientific activities: normal and revolutionary; for her 'there is just normal science, understood as standard testing' (p. 55). This criticism of Kuhn is most welcome: it retains the principal features of what Kuhn calls normal science, but recasts this practice into a different role. While for Kuhn anomalies give rise to research puzzles, for Mayo anomalies afford opportunities for learning—learning from error (p. 55). Mayo seeks then to accommodate a more realistic and less theory-dominated picture of inquiry. She achieves that by divorcing normal, standard testing from the Kuhnian dependence upon background paradigms in any sense other than dependence upon an intertheoretic pool of exemplary models of error. She observes sensibly that 'in much of day-to-day scientific practice, and in the startling new discoveries we read about, scientists are just trying to find things out' (p. 56). Following her realistic view of the practice of science, Mayo recommends going smaller, not bigger—to the local tests of 'normal science' (p. 57). In contrast to the thrust of holistic models, Mayo advocates looking to the force of low-level methods of experiment and inference.

> The fact that theory testing depends on intermediate theories of data, instruments, and experiment, and that data are theory laden, inexact, and 'noisy,' only underscores the necessity for numerous local experiments, shrewdly interconnected (p. 58).

Mayo's well founded criticism of Popper, Kuhn and, in general, of all holistic models lands her squarely in the camp of the New Experimentalists. She finds herself in respected company: Robert Ackermann, Nancy Cartwright, Allan Franklin, Peter Galison, Ronald Giere, and Ian Hacking (p. 58). Although their agendas, methods, and conclusions differ, they share the core thesis that

> focusing on aspects of experiment holds the key to avoiding or solving a number of problems, problems thought to stem from the tendency to view science from theory-dominated stances (p. 58).

Traditionally, theories of confirmation, inductive inference, and testing were conceived of in a theory-dominated philosophy of science; the New Experimentalists wish to move away from this approach. For that purpose,

> [t]he old-style accounts of how observation provides an objective basis for appraisal *via* confirmation theory or inductive logic should be replaced by an account that reflects how experimental knowledge is actually arrived at and how it functions in science (p. 60).

As Mayo explains, the New Experimentalists wish to sidestep the philosophical paradoxes and difficulties that plagued formal attempts at inductive logic. The complexities and context dependencies of actual experimental practice just seem recalcitrant to the kind of uniform treatment dreamt of by philosophers of induction (p. 67). Mayo approvingly quotes Galison who emphasizes that

> experimentalists' real concern is not with global changes of world view. In the laboratory the scientist wants to find local methods to eliminate or at least quantify backgrounds, to understand where the signal is being lost, and to correct systematic errors (p. 61).

This is by now a well-trodden path. But Mayo continues with her incessant critical crusade and takes the reader further afield. She rightly remarks that no systematic programme has emerged from the understanding that local narratives about experiments offer a rich source from which to extricate information on how reliable data are obtained and used for learning about experimental processes. Narratives are not enough (pp. xi, 12). It is necessary to get at the structure of experimental activities and at the epistemological rationale for inferences based on such activities (p. 58).

This is a crucial juncture in the development of the book—the point, so to speak, of no return. In explaining why the New Experimentalists have come up short, Mayo prepares the ground for her philosophy of experiment. The New Experimentalists, she maintains, have not yet tapped those aspects of experiment that provide the tools for understanding experimental activities and their epistemological rationale. These aspects

> cover the designing, modeling, and analyzing of experiments, activities that receive structure by means of statistical methods and arguments (p. 58).

To be sure, Mayo is aware that many a narrative of experiment is replete with applications of statistical techniques to arriving at data, to assessing the fit of data to a model, and to distinguishing real effects from artifacts (e.g. techniques of data analysis, significance tests, standard errors of estimates, and other methods from standard error statistics) (pp. 58–59). Indeed,

> [t]he experimental narratives themselves are chock-full of applications of standard statistical methods, methods developed by Fisher, Neyman and Pearson, and others (p. 68).

What has not been done, however, she argues,

> is to develop these tools into something like an adequate philosophy or epistemology of experiment. What are needed are forward-looking tools for arriving at reliable data and using such data to learn about experimental processes (p. 59).

Despite the alleged commitment of the New Experimentalists to the actual practices of science, there has not been an attempt to explicate statistical practices in scientists' own terms (p. 68). Mayo therefore encourages the New Experimentalists to overcome their resistance to employing statistical ideas in setting out a general account of experimental inference. She frowns on those New Experimentalists who ironically revert to the theory-dominated philosophy of decision and inference, particularly of the Bayesian kind, when they employ formal statistical ideas to give an overarching structure to experiment (pp. 68–69).

> The proper role for statistical methods in an adequate epistemology of experiment... is not the theory-dominated one of reconstructing episodes of theory confirmation or large-scale theory change. Rather their role is to provide forward-looking, ampliative rules for generating, analyzing, and learning from data in a reliable and intersubjective manner (p. 59).

The New Experimentalism may reveal the function and rationale of statistical tools from the perspective of actual experimental practice—the very understanding which is missing from theory-dominated perspectives of scientific inference such as Bayesianism. This understanding is the basis for both Mayo's critique of the Bayesian Way and for her defense of standard error statistics. 'The Bayesian Way is the wrong way to go,' (p. 59) she declares and presses on with her own programme of New Experimentalism: an error-statistical philosophy of experiment.

It should be noted that Mayo's pungent criticism of the Bayesian Way is directed solely to the application of Bayesian inference to philosophical problems of scientific inference and hypothesis testing. She rightly distinguishes this context from other statistical contexts such as personal decision-making, where the application of Bayesian inference requires ingredients (e.g. prior probabilities) less problematic than those found in the context of philosophy of science (p. 69). Notwithstanding, Mayo's critique of the application of the Bayesian inference to issues in philosophy of science in general and to experimentation in particular, undoubtedly poses a great challenge to the adherents of this inference. If they wished to defend their Bayesian stance against Mayo's criticism, they would have to think of new arguments.

Mayo's philosophy of experiment bridges two different views: the view that the key to solving problems in philosophy of science is an inductive–statistical account of hypothesis appraisal and the view that the key is an understanding of the nature and role of experiment in scientific practice. The bridge Mayo builds is designed then to connect, as she puts it, the cornerstone of logical empiricism with the centerpiece of the New Experimentalism (p. 126). Essentially, Peircean induction and Neyman–Pearson statistics (p. 413) provide the foundation at the one end of the bridge while the foundation at the other end is provided by Suppean hierarchy of models, models for analysing the relation between empirical theories and the relevant data (p. 132).

Following Patrick Suppes (pp. xi, 130–131), Mayo proposes to view experimental inquiry in terms of a hierarchy of models ranging from primary scientific hypotheses to the nitty-gritty details of the generation and analysis of data. She argues in the spirit of the New Experimentalism that an adequate account of experimental testing should not begin at the point where data and hypotheses are simply given. One cannot just throw some 'evidence' at the error statistician and expect an informative answer to the question of how well it warrants a hypothesis. Rather, one must explicitly incorporate the intermediate theories of data, instruments, and experiment that are required to obtain experimental evidence (pp. 128, 444).

The Suppean idea is to conceive of experiment as consisting of a hierarchy of models: primary models, experimental models and data models (Chapter 5). The hierarchy allows not only for assessing the relation between empirical theories and the relevant data, but also for addressing systematically the key questions of an epistemology of experiment: questions about what data to collect, how to model them, how to check their assumptions, how to use them to learn about experimental processes, and how to relate experimental knowledge to scientific hypotheses (pp. 128, 444–445). Mayo's chief interest is in spelling out and analyzing the relationships between the data models and the primary models (p. 131).

The Suppean hierarchy of models is linked as it were on the Mayo bridge to Peircean induction and Neyman–Pearson (NP) statistics. In a nutshell, the NP testing defines, as Mayo explains,

> mathematical functions on random variables. The variables may take on different values corresponding to different outcomes of an experiment. Tests are functions that map possible values of these variables (i.e., possible experimental outcomes) to various hypotheses about the population from which outcomes may have originated (p. 365).

The test maps each of the possible outcomes—the experimental *sample space*—onto either *H* or *J* (hypothesis *H* against alternative *J*); those mapped onto *H* (i.e., into 'accepting' *H*) form the *acceptance region*, those mapped onto alternative *J* form the *rejection* (*of H*) *region*. The model which Mayo advocates describes the partitioning that results from the mapping rules. The tests focus on the probabilistic properties of these mapping rules, that is, on the probabilities with which the rule leads to one or another hypothesis, under varying assumptions about the true hypothesis (p. 366). The standard terminology is to associate these alternatives with two types of probabilistic error: error of the first kind (type I) and error of the second kind (type II). Mayo's central thesis is that the argument from error, that is, learning from error, may be described in terms of a test of a hypothesis, *H*, that a given error is absent. The evidence indicates the correctness of hypothesis *H*, when *H* passes a severe test—one with a high probability of failing *H*, if *H* is false. An analogous argument is used to infer the presence of an error (p. 64). This then is the framework of Mayo's error-statistical philosophy of experiment.

Mayo's philosophy of experiment relies neither on scientific theories nor on a theory of experiment; rather, it relies on methods—statistical methods—for producing experimental effects (p. 15). This observation, I suggest, is crucial. It explains the limited view of experiment exhibited in this study. In spite of the fact that Mayo speaks voluminously about the need to address the actual practice of experimentation, she focuses her attention solely on statistical calculations. As Mirowski, criticizing both Bayesians and classical statisticians, aptly puts it: 'the empirical inquirer cranks through the formulas, assigns the error probabilities and reports an outcome—all as a hermetically self-contained procedure' (1995, p. 542). This is not what one would expect of, say, a Faraday, a Helmholtz, a Pasteur, a Hertz, a Rutherford, a Gibson, a Rabi or a Kapitza. Consider Peirce's observations on experimental style:

> Of all men of the century Faraday had the greatest power of drawing ideas straight out of his experiments and making his physical apparatus do his thinking, so that experimentation and inference were not two proceedings, but one. To understand what this means, read his *Researches on Electricity*. His genius was thus higher than that of Helmholtz, who fitted a phenomenon with an appropriate conception out of his store, as one might fit a bottle with a stopper (Peirce, 1966, p. 272).

Mayo's 'full-bodied experimental philosophy' (p. 444) is not attuned to the act of experimenting; it focuses rather on the end result: data and their statistical tests. Questions of interpretation, for example, do not arise in this frame-work (for a different approach see Radder, 1995; Hon, 1989a, b, 1998).

Mayo is not oblivious to this difficulty. She observes that 'checking assumptions does not always call for running explicit statistical tests'. She admits that she is 'only scratching the surface here. A full-blown philosophy of experiment calls for much more work in explicating the formal tests and the informal reasonings that go into checking experimental assumptions' (p. 138; cf. 434–435). This is indeed the case, but in her classical testing approach, every experimental situation is rendered statistical.

> Because of the many sources of approximation and error that enter into arriving at the data, the data would rarely if ever be expected to agree exactly with theoretical predictions. As such, the link between data model and experimental hypothesis or question may often be modeled statistically, *whether or not* the primary theory or hypothesis is statistical. This statistical link can be modeled in two ways: the experimental prediction can itself be framed as a statistical hypothesis, or the statistical considerations may be seen to be introduced by the test (in its using a statistical test rule) (p. 134; emphasis added).

Notwithstanding, Mayo admits that her philosophy of experiment is limited: 'the statistical theory of experiment deals *only with certain kinds of experiments* insofar as their behavior may be characterized by certain parameters. A characteristic of key interest is the relative frequency with which an outcome occurs, or would occur, in a sequence of applications of the experiment in question' (pp. 161–162; emphasis added; cf. pp. 164, 173). Hence the importance of finding ways of *extruding* the data from various models to provide a result for '*what it would be like*'... (pp. 363, 430, 445, 459, 461). 'Virtually all the uses of statistical ideas in learning from error throughout this book,' Mayo states,

> depend critically on such considerations of 'would have beens.' What makes standard error statistical tools so useful for scientific inference is that their formal properties, error probabilities, enable learning about what would be expected if various errors exist—the key to experimental arguments from error (p. 358).

Mayo does not use the term 'extrusion;' she does however use the term 'massaging.' Massaging the data is a standard practice that should not be frowned upon; still, it should be borne in mind and weighed (pp. 15, 226). This approach leads 'naturally' to the idea of doing experiments 'on paper' (p. 430). Experimental arguments, Mayo suggests, often serve as surrogates for actual experiments; they may be seen as experiments 'done on paper'. Running tests and simulations on computers is the next 'natural' step (p. 447). It is therefore no surprise to the reader that for Mayo applying statistical methods may be seen, in a very real sense, as continuing experimentation by other means (mimicking van Fraassen paraphrasing Clausewitz) (p. 459; cf. the motto of this review).

Mayo does not see a difference between a hierarchy of models and a flow chart of experiment; 'nothing turns on,' she remarks, 'whether we choose to array the models of inquiry into a hierarchy or into something like the flow chart' (p. 133). The author opts without hesitation for the Suppean hierarchy of models. But this very hierarchical structure allows her to compartmentalize the experiment, concentrate on one aspect of it, that is, the data modeling, and sever it from the rest of the act of experimenting (for a flow-chart view of experiment see Hon, 1989a, p. 480; cf. Gooding, 1990). In Mayo's philosophical framework, experimental knowledge becomes completely statistical:

> experimental knowledge is knowledge of the probabilities of specified outcomes in some actual or hypothetical series of experiments. Its formal statement may be given by an *experimental distribution* (a list of outcomes and their associated probabilities), or by a standard 'random' process such as a coin-tossing mechanism (p. 12, original emphasis; cf. pp. 162, 461).

Success in obtaining experimental knowledge is explained, according to Mayo, by the properties of the statistical methods applied.

> Because we can frame questions of interest in term of hypotheses amenable to severe testing, we are able to learn from error and in so doing obtain experimental knowledge (p. 441).

The properties of the methods are in fact, as Mayo states, error probabilities (p. 441) which permits quantifying trustworthiness (pp. 424, 425). Thus the errors upon which Mayo builds her error-statistical philosophy of experiment are not error at large but rather a specific and indeed limited kind of error, namely, error probabilities. Error probabilities are not probabilities of hypotheses, but the probabilities that certain experimental results would occur, were one or another hypothesis true about the experimental system (p. 367).

Error, as I have remarked, covers multiple sins. What kind of error did Franck and Hertz commit in their Nobel winning experiment? What happened in Ehrenhaft's experiments which made him conclude that there are subelectrons? What went wrong in Kaufmann's experiments so that he could speak decisively against the correctness of Einstein's special theory of relativity? How did Blondlot discover the existence of a new form of radiation—N rays—which does not exist? Why was Lowell so convinced that the visible lines on the surface of Mars are in fact artificial canals for irrigation purposes? 'The history of science,' Maxwell observed, 'is not restricted to the enumeration of successful investigation. It has to tell of unsuccessful inquiries, and to explain why some of the ablest men have failed to find the key of knowledge.' (Maxwell, 1871, p. 251). The probabilistic approach to the study of error is undoubtedly of considerable importance; of no less importance is the study of conceptual and physical circumstances in which errors in experimentation may originate. 'One must,' as Wittgenstein demanded, 'reveal the source of error' (Wittgenstein, 1979, p. 61).

While Mayo's two-pronged plan—obstructing the Bayesian Way and establishing an error-statistical philosophy of experiment—is laudable, the philosophical gains should be questioned. Reflecting upon the title of the book, questions arise as to what is the nature of the errors upon which Mayo builds her programme and, of equal importance, what is the nature of the knowledge she characterizes as experimental. The errors, as we have seen, are of a particular kind; they are error probabilities. Furthermore, the insight claimed to have been obtained in the framework of this philosophy into experimental knowledge is, again, of a very particular kind; it is in fact an insight into statistical methods of assessing data. These methods are, to be sure, of crucial importance in experimentation but they do not comprise in themselves a philosophy of experiment.

What we have then as the theme of the book is 'Error Probabilities and the Statistical Assessment of Experimental Data.' Thus, while Mayo's book is rich in philosophical reflections which will no doubt be of much use to many a philosopher of science, it constitutes nevertheless a contribution to experimental design (in the traditional sense of the term) as well as to analysis of error probabilities—the very subject which is commonly classified as QA275. It transpires then that the librarian of the Library of Congress has been in the right all along.

Mayo has mapped philosophical issues concerning error probabilities and statistical testing of experimental data. Charting these specific aspects of error and experiment has made conspicuous the need for systematic explorations of three major subjects of research in this field: (1) a study of the history of error in science generally and in experimentation in particular[3] and (2) an epistemology of experiment that can inform a history of error *via* (3) a classification of types of error that reflects this epistemology.[4] Errors should be exploited, but there is still much prospecting to do before we can fully benefit from the knowledge that can be extracted from this widespread epistemological phenomenon.

Acknowledgements—I thank Oded Balaban, Ruth Lorand, Menachem Luz, Shaul Smilansky and Zeno G. Swijtink for advice and valuable comments.

References

Bates, D. W. (1996) 'The Epistemology of Error in Late Enlightenment France', *Eighteenth-Century Studies* **29,** 307–327.

Bremer, J. M. (1969) *Hamartia: Tragic Error in the Poetics of Aristotle and in Greek Tragedy* (Amsterdam: Hakkert).

Buchwald, J. Z. (1995a) 'Why Hertz was Right about Cathode Rays', in Buchwald (ed.) (1995b), pp. 151-169.

Buchwald, J. Z. (ed.) (1995b) *Scientific Practice: Theories and Stories of Doing Physics* (Chicago: Chicago University Press).

Cornford, F. M. (1957) *Plato's Theory of Knowledge* (New York: The Liberal Arts Press).

Feigl, H. and Maxwell, G. (eds) (1961) *Current Issues in the Philosophy of Science* (New York: Holt, Rinehart and Winston).

Galison, P. (1987) *How Experiments End* (Chicago: University of Chicago Press).

Gooding, D. (1990) *Experiment and The Making of Meaning* (Dordrecht: Kluwer Academic Publishers).

Hacking, I. (1965) *Logic of Statistical Inference* (Cambridge: Cambridge University Press).

Heidelberger, M. and Steinle, F. (eds) (1998) *Experimental Essays—Versuche zum Experiment* (Baden-Baden: Nomos).

Hon, G. (1987a) 'H. Hertz: "The Electrostatic and Electromagnetic Properties of the Cathode Rays are either *Nil* or very Feeble". (1883) A Case-Study of Experimental Error', *Studies in History and Philosophy of Science* **18,** 367–382.

Hon, G. (1987b) 'On Kepler's Awareness of the Problem of Experimental Error', *Annals of Science* **44,** 545–591.

Hon, G. (1989a) 'Towards a Typology of Experimental Error: an Epistemological View', *Studies in History and Philosophy of Science* **20,** 469–504.

Hon, G. (1989b) 'Franck and Hertz versus Townsend: a Study of Two Types of Experimental Error', *Historical Studies in the Physical and Biological Sciences* **20,** 79–106.

Hon, G. (1989c) 'Is There a Concept of Experimental Error in Greek Astronomy?', *British Journal for the History of Science* **22,** 129–150.

[3]See e.g. Hon (1987b, 1989c); Krüger *et al.* (1989); Mirowski (1994); Sheynin (1983); Stigler (1986) and Swijtink (forthcoming). For further references see Mirowski (1995, p. 546, note 8).

[4]See e.g. Gooding (1990); Radder (1995); Hon (1989a, b, 1995, 1998).

Hon, G. (1995) 'Going Wrong: to Make a Mistake, to Fall into an Error', *Review of Metaphysics* **49,** 3–20.
Hon, G. (1998) 'If This Be Error: Probing Experiment With Error', in Heidelberger and Steinle (eds) (1998).
Koyré, A. (1945) *Discovering Plato* (New York: Columbia University Press).
Krüger, L., Daston, L. J. and Heidelberger, M. (eds) (1989) *The Probabilistic Revolution: Ideas in History* (Cambridge, MA and London: The MIT Press), 2nd printing.
Lattimore, R. (1965) *The Iliad of Homer* (Chicago and London: Phoenix Books).
Lewis, C. T. and Short, C. (1966) *A Latin Dictionary* (Oxford: Clarendon Press).
Maxwell, J. C. (1871) 'Introductory Lecture on Experimental Physics', in his (1890) *Scientific Papers*, ed. W. D. Niven, Vol. II (Cambridge: Cambridge University Press), pp. 241–255.
Mellor, D.H. (1967) 'Imprecision and Explanation', *Philosophy of Science* **34,** 1–9.
Mirowski, P. (1994) 'A Visible Hand in the Marketplace of Ideas: Precision Measurement as Arbitrage', *Science in Context* **7,** 563–589.
Mirowski, P. (1995) 'Civilization and Its Discounts', *Dialogue* **34,** 541–559.
Peirce, C. S. (1966) *Selected Writings*, edited with an introduction and notes by P. P. Wiener (New York: Dover).
Popper, K. (1974 [1963]) *Conjectures and Refutations: the Growth of Scientific Knowledge* (London: Routledge and Kegan Paul), 5th edition.
Quine, W. V. (1995) *From Stimulus to Science* (Cambridge: Harvard University Press).
Radder, H. (1995) 'Experimenting in the Natural Sciences: a Philosophical Approach', in Buchwald (ed.) (1995b), pp. 56–86.
Sellars, W. (1961) 'The Language of Theories', in Feigl and Maxwell (eds) (1961), pp. 57–77.
Shapin, S. (1994) *A Social History of Truth* (Chicago, IL: University of Chicago Press).
Sheynin, O.B. (1983) 'Corrections and Short Notes on My Papers', *Archive for the History of Exact Sciences* **28,** 171–195.
Stigler, S. M. (1986) *The History of Statistics: the Measurement of Uncertainty Before 1900* (Cambridge, MA and London: The Belknap Press of Harvard University Press).
Swijtink, Z. G. (forthcoming) 'Error Theory', in A. Hessenbruch (ed.), *Reader's Guide to the History of Science* (London: Fitzroy Dearborn).
Wittgenstein, L. (1979) 'Remarks on Frazer's *Golden Bough*', in C. G. Luckhardt (ed.) *Wittgenstein Sources and Perspective* (Hassocks, Sussex: Harvester Press), pp. 61–81.