REVIEW

DEBORAH MAYO
Error and the Growth of Experimental Knowledge
Chicago and London, The University of Chicago Press, 1996

Hasok Chang
Department of Science and Technology Studies
University College London

Deborah Mayo's *Error and the Growth of Experimental Knowledge* is a rich, useful, and accessible book. It is also a large volume which few people can realistically be expected to read cover to cover. Considering those factors, the main focus of this review will be on providing various potential readers with guidelines for making the best use of the book.

As the author herself advises, the main points can be grasped by reading the first and the last chapters. The real benefit, however, would only come from studying some of the intervening chapters closely. Below I will offer comments on several of the major strands that can be teased apart, though they are found rightly intertwined in the book.

Scientific change and progress. One laudable characteristic I find throughout the volume is that it grapples with the traditional questions seriously, instead of abandoning them as passé. From that point of view I can heartily recommend Chapter 2, 'Ducks, Rabbits, and Normal Science: Recasting the Kuhn's-Eye View of Popper'. This chapter can be read quite well even in isolation from the rest of the book. I think Mayo's insightful synthesis of Popper and Kuhn will be useful to everyone, from advanced undergraduates to experts in the field. The basic idea is that Kuhnian normal science, when interpreted as Mayo proposes, 'turns out to offer an effective basis for severe testing' (p. 23), quite contrary to the fear of dogmatism that Popper expressed concerning normal science. The only reservation I have is the feeling that Kuhn's ideas get flattened. Particularly, the notion of incommensurability is almost completely ignored, to the point where Mayo rejects the notion of extraordinary or revolutionary science altogether—'there is just normal science, understood as standard testing' (p. 55). To be fair, I should note that Mayo does not regard it as her task to give a faithful account of 'what Kuhn saw himself as doing' (p. 21).

Learning from error. The most fundamental tenet of Mayo's philosophy, in agreement with Popper, is that we learn from error, and she seeks to describe

and recommend concrete methods for doing that. The basic ideas can be grasped from Chapters 1 and 5. In a way, the overall direction of Mayo's arguments could be characterized as a return to scientific common sense. A heavy emphasis is placed on the measure of statistical significance and other related concepts, whose widespread application in contemporary scientific practice cannot be denied.

I think more reflections on what 'error' means would be helpful for the entire project. On the one hand, Mayo's typology of error (first given on p. 18) could be developed and elaborated further with much benefit. On the other hand, Mayo's use of the word 'error' seems to indicate perhaps an overly straightforward sense of a result that deviates from the truth. I do not mean to imply that this book should have included lengthy discussions mired in references to the existing literature on truth and realism. Still, talking about 'learning from error' involves the assumption that we can identify an error when we see one, and know when we are getting closer to the truth. (So it is not a surprise that when it comes to Kuhn, only normal science can be nicely incorporated into Mayo's way of thinking.)

Mayo's operative tools derive from standard Neyman–Pearson statistics. Starting from a specification of a population, we can compute the probabilities of error inherent in reaching conclusions on the basis of given methods of sampling and inference. The situation is not quite the same in empirical investigations, in which we neither know the true distribution to begin with, nor have any guarantee that the sampling is actually done as we intend it; so our estimates of error probabilities must rely on assumptions that are ultimately unfounded, or at least unverified. But these worries are fully anticipated in Mayo's discussions, as explained below.

Self-correction and ampliative inference. The part of Mayo's work that I find most exciting is the argument that scientific knowledge can transcend the problem of induction (and not obviate it, as Popper would have it). Drawing on Peirce's ideas (Chapter 12), she argues that well-known experimental and statistical techniques allow us to spot and correct our errors, even without an ultimate and direct access to the truth. Moreover, inductive inferences can be truly ampliative, meaning that we can properly reach conclusions that are more informative than the data would logically allow. Concrete details illustrating these two major points are scattered throughout the volume, but something of a summary can be found in Chapter 13. I think much more work will be needed to develop and defend these points, and I hope Mayo's work will continue to stimulate further enquiries.

Levels in testing. One crucial ingredient in Mayo's picture of scientific knowledge is the hierarchy of models or, more generally, different levels of knowledge. This is spelled out in great detail in Chapter 5, but an examination of Table 5.1 (p. 130) and Table 5.2 (p. 140; inspired by Suppes) may be sufficient for readers

who are familiar with related ideas from other authors, whether they be Feigl (empirical laws vs. high-level theory), Cartwright (phenomenological vs. fundamental laws), Bogen and Woodward (data vs. phenomena), or indeed Suppes. At any rate, Mayo's notion is that learning from error only becomes manageable if we can break down enquiry into smaller bits; trying to go from raw data to major theoretical hypotheses will only invite underdetermination, seemingly insurmountable.

In this context, Mayo gives a very informative discussion of two well-known historical cases: Perrin's work on Brownian motion, and the eclipse expeditions to test Einstein's prediction of the bending of starlight around the sun. Both Perrin and Eddington are seen as accepting or rejecting their data on the basis of low-level assumptions that can be tested. As described by Mayo, Perrin (see Chapter 7) is very convincing in demonstrating the 'random walk' nature of Brownian motion. On the other hand, the arguments attributed to him are pretty weak when it comes to arguing for the molecular-kinetic theory in general. The situation is similar in the other case (Section 8.6): Eddington seems to have convincing error-probability arguments for his conclusions about the amount of the observed deflection; on the other hand, when it comes to understanding how this result was used as evidence for general relativity, there doesn't seem to be very much that the error-probability framework can give us.

Severity. In the concrete discussions of historical cases and methodological quarrels, the centre-piece of Mayo's reasoning is the criterion of severity; this is spelled out in detail in Chapter 6 (especially pp. 178–84). To put it informally, a severe testing procedure has a high probability of rejecting a hypothesis that is incorrect; passing a severe test confers credibility to a hypothesis. Mayo is not operating in a falsificationist context, and passing a severe test actually confers reliability to a hypothesis, not just Popperian corroboration. This is plausible provided that we accept Mayo's confidence about the low-level assumptions which enter the calculation of severity (see 'Levels in testing' above). A wide array of methodological intuitions and scientific decisions receive justification as moves to increase severity in testing, and as instances of accepting the verdict of severe tests (see 'Strategies of testing' below).

Bayesianism. Attacks on what Mayo calls 'the Bayesian Way' can be found in almost all of the discussions contained in the book. The central points can be gathered from Sections 3.3 and 3.4, and Chapter 10. I think Mayo is right in predicting that she will not convince the faithful Bayesians, but I think she does manage to launch a stimulating attack on the 'only game in town'. I hope some Bayesians will respond constructively to various aspects of Mayo's critique and polemics, to carry the debate forward.

I think the most notable aspect of Mayo's stance is her steadfast refusal to attach probabilities to hypotheses. Such probabilities would really make sense

only if 'universes were as plenty as blackberries' (Peirce's phrase) so that we could consider the number of universes in which the given hypotheses are true. This world-counting is not necessary for making sense of error probabilities, which are just relative frequencies with which testing procedures would produce erroneous verdicts about hypotheses. This will strike many critics as hard-headed frequentism, but I think there is good sense in trying to see how far we can get just on relative frequencies, which after all must be the least problematic way of understanding and handling empirical probabilities. Readers can judge for themselves the fruitfulness of Mayo's frequentist approach by examining the positive results produced, rather than focusing on the argument with Bayesians.

Strategies of testing. There is much valuable detailed discussion of testing strategies in the latter half of the book. I think the pivotal point in these discussions is that data should be considered not only in themselves but also in relation to other sets of data that would be produced by applying the same testing procedure in similar situations. This comes to the point made above (see 'Bayesianism'), that we should talk about the reliability of testing procedures rather than the degree of belief or confirmation conferred on hypotheses by data.

There are three major issues I would like to highlight from Mayo's discussions. First, Mayo grapples (in Chapter 8) with the old question regarding the value of novel evidence in confirmation: does a hypothesis receive more confirmation when it predicts novel phenomena that later get verified, than when it merely accommodates phenomena that are already known? Building on and adding to the excellent treatment of this issue by Musgrave, Mayo argues convincingly that it is severity that matters, not whether the evidence was known or used in the construction of the hypothesis. I am, however, not entirely convinced by her criticism of the views advanced by Worrall and Giere, and I think it would be fruitful to solicit responses from these authors.

On the other two issues, Mayo comes out more explicitly against the Bayesians concerning their stance that the only thing that matters is the actual data, not how the data were obtained. Mayo argues against the strategy of 'hunting' for correlations (Chapter 9) and the 'try and try again' strategy for getting the result that one would like to see (Section 10.3), because these testing procedures lack severity. For the Bayesians the lack of severity here is imaginary (to the extent that they can conceptualize severity as construed by Mayo), because what determines the posterior probability is the evidence itself (together with the prior probabilities), not the evidence plus what else would have been obtained by applying the same procedure again.

On these questions of testing strategies, I think it would be very informative to generate more comparative discussions on how actual scientific decisions would be (or would have been) made differently depending on which philosophy

or method of testing one adopts. It is not as if the Bayesian probability kinematics could produce decisive verdicts on many actual scientific issues, and a similar thing can probably be said about the error-statistical approach when it comes to the high-level scientific decisions that methodologists like to discuss (see 'Levels in testing' above). My own suspicion is that disagreements between the methods would occur mostly in cases that are so uncertain that no consensus would be reached even if only one method is used.

There is much else in Mayo's book that I do not have the space to discuss here. I will briefly note that there is a good deal of discussion of Duhem and underdetermination (Chapters 4 and 6), which I personally found less useful than the discussions highlighted above. Mayo's remarks on Neyman's and Pearson's differing philosophies (Chapter 11) and Peirce's views on inductive inference (Chapter 12) will be of interest to the appropriate specialists, and they also do give further insights into Mayo's own ideas.

In sum, I hope many readers will make use of this major work and carry on the debates it aims to stimulate.