Error and the Law
Exchanges with Larry Laudan

Deborah G. Mayo

As with each of the contributions to this volume, my remarks on Larry Laudan reflect numerous exchanges over a long period, in this case, since he was a colleague in the 1980s. Here, we put to one side the discussions we have had on matters of theory testing (although they arise in previous chapters) and shift to extending our exchanges to a new project Laudan has thrown himself into over the past decade: legal epistemology. Laudan's idea of developing a legal epistemology takes us to the issue of how philosophers of science may develop accounts of evidence relevant to practice. I heartily concur with Laudan's invitation to philosophers of science to develop a "legal epistemology" and agree further that such an "applied epistemology" calls for a recognition of the role of burdens of proof and the need to "have a self-correction device for identifying and *revising our erroneous beliefs*" (Laudan, this volume, p. 376). My goal will be to point out some gaps that seem to need filling if one is to clear the way for this interesting program to succeed. Some queries include:

1. *Experimental Reasoning and Reliability*: Does/should probability enter in distinct ways in standards of evidence in law? (e.g., do the standards BARD and POE differ by degree or by kind?)
2. *Objectivity and Rationality*: Do "standards of evidence" vary across sciences? Do differing assessments of "costs" of mistaken inferences preclude a single standard for determining what conclusions are warranted from evidence? Does the latitude in setting standards of evidence preclude viewing scientific inference as a unified, objective, rule-governed activity?
3. *Metaphilosophical concerns*: What are the responsibilities of the "two-way street" between philosophy and practice?

Laudan's discussion offers an excellent example of how "applied philosophy of science" involves us in a two-way street: (a) philosophical analysis can point out conceptual and logical puzzles that often go unattended in practice; at the same time, (b) studying evidence in practice reveals problems (and also solutions) often overlooked in philosophy. Under (a), the most basic issue that cries out for clarity concerns the very meaning of the terms used in expressing these standards of evidence: evidence beyond a reasonable doubt (BARD) and that of the preponderance of evidence (POE). This aspect of the work is of a traditional philosophical kind – alleviating the conceptual puzzles of others – but in a new arena, which has the added appeal of requiring astuteness regarding probabilistic concepts in evidential standards. Going in the other direction, (b), Laudan's discussion leads us to consider whether variability of standards in legal practice challenges "the conviction that science is a quintessentially rational, rule-governed activity" (p. 395). His provocative argument suggests the answer is no.

1 Explicating Standards of Proof in the Law: SoPs

Laudan plausibly links the two familiar standards of evidence – beyond a reasonable doubt (BARD) and the preponderance of evidence (POE) – to that of controlling and balancing the rates of error, much as in conducting statistical tests.

> With BARD we expect a low rate of false positives, and we are willing to absorb a relatively high rate of false negatives, if necessary, in order to keep false convictions to an acceptable level. By contrast, the preponderance of the evidence standard implicitly but unequivocally denies that one sort of error is more egregious than the other. (Laudan, this volume, p. 383)

1.1 Interpreting BARD

Let us begin with the BARD requirement. The rationale in law for ensuring a small probability of erroneous convictions (false positives) is akin to the intended use of controlling the type I error in statistics. In statistics, one is to choose the null hypothesis, H_0, so that erroneously rejecting H_0 is the error that is "first" in importance (even though this is not always adhered to); hence, it is the "type I error" (for a discussion, see Mayo and Spanos, 2006). Failing to reject the null hypothesis with data **x** is akin to failing to convict a defendant on the evidence brought to trial. By fixing at a low value the probability of rejecting the null when it is true (i.e., committing a type I error), the null hypothesis, like the defendant's innocence, is being protected. Within that stipulation, as in statistical tests, one seeks the test

with the greatest power to detect that the null hypothesis is false (i.e., the one with the smallest probability of failing to reject the null when it is false, the type II error).

In Laudan's recent, excellent book (Laudan, 2008), the analogy with statistical tests (using approximately normal distributions) is more explicit than in this chapter. Just as standard statistical tests may be seen to consist of accept/reject rules, a trial may be seen as a rule that takes (the sum total of) evidence into either one of two verdicts: acquit or convict a defendant. That is, a trial uses evidence to infer either

H: innocence (i.e., defendant comes from the innocent population)
or
J: guilt (i.e., the defendant comes from the guilty population).

The rule is determined by specifying the extent of evidence required to reach one or the other verdicts, which in turn is to reflect chosen rates for erroneous convictions (type I error) and erroneous acquittals (type II error).

Something like a ten-to-one ratio of type II to type I error probabilities, Laudan explains, is commonly regarded as the approximate upshot of BARD: "Unless we think that the social costs of a false conviction are roughly ten times greater than the costs of a false acquittal, then we have no business setting the standard of proof as high as we do." By contrast, Laudan goes on to say, "If, for instance, we regarded the two sorts of mistakes as roughly equally costly, a preponderance standard would obviously be the appropriate one because it shows no bias toward one sort of error over the other." (Laudan, this volume, p. 384)

Type I error probability: P(test *T* leads to conviction; *H*-defendant is innocent) = α (fixed at a small value, say .01).[1]

Type II error probability: P(test *T* leads to acquittal; *J*-defendant is guilty) = β (set at, say, 10 times α (i.e, around .1); the power of the test is then .9).

Note that the error probability attaches to the generic event that the trial outputs evidence that reaches the SoP that is set – not to a specific set of evidence. Setting the SoP to be BARD is akin to requiring the evidence *e* to be statistically significant at the .01 level – which may be abbreviated as SS(.01). We get:

P(*e* is SS(.01); *H*-defendant is innocent) = .01.

---

[1] This may be read: The probability that the test (trial) convicts the defendant, under the assumption that she is innocent, equals α.

We may understand "*e* is SS(.01)" in this context as the event that the evidence is sufficiently "far" from what would be expected under the null hypothesis (innocence) that such evidence would be expected no more than 1% of the time were the evidence to have truly come from an innocent person. The type II error rate, accordingly, is

P(*e* is not SS(.01); *J*-defendant is guilty) = .1.

1.2 Interpreting POE

Laudan regards a "preponderance standard," POE, by contrast to BARD, as holding "the two sorts of mistakes as roughly equally costly" and so, presumably, $\alpha = \beta$, although he does not say how small each should be. Should they both be .1? .01? .05? Setting $\alpha = \beta$ does not ensure both are small, nor does it preclude their being set even smaller than .01. So already we see a need for clarification to make out his position.

What Laudan is clear about is that he regards BARD and POE as merely differing by degree, so that whatever levels of evidence they require, they are not measuring different things. They refer to the same "bar," just raised to different heights. Some of his remarks, however, suggest slippage between other potential construals of POE. For example, Laudan also construes finding a POE for some hypothesis as evidence that the hypothesis is "more likely than not." In particular, in his "full-blown version" of an affirmative defense (AD), the defendant asserting a given defense must persuade the jury that his AD is more likely than not (pp. 379, 381).

Let H_{AD} be the hypothesis that the affirmative defense holds true. Then this "more likely than not" assertion, employing the statistical notion of likelihood, asserts the following:

$\mathrm{Lik}(H_{AD}; \boldsymbol{e}) > \mathrm{Lik}(\text{not-}H_{AD}; \boldsymbol{e})$

where $\mathrm{Lik}(H_{AD}; \boldsymbol{e})$ abbreviates the likelihood of H_{AD} given $\boldsymbol{e}$, which is defined as

$P(\boldsymbol{e}; H_{AD})$

(see Chapter 7). It is important to see that in calculating likelihoods the evidence $\boldsymbol{e}$ is fixed – it refers to a specific evidence set; what varies are the possible hypotheses under consideration. So to say that H_{AD} is more likely than not is to say

(i) $P(\boldsymbol{e}; H_{AD}) > P(\boldsymbol{e}; \text{not-}H_{AD})$.

The likelihood ratio is best seen as a "fit" measure, so that statement (i) says that H_{AD} fits e better than not-H_{AD} does. However, this does not entail anything about error probabilities α and β! So, perhaps the initial impression that Laudan regards both BARD and POE as error-probabilistic standards, differing only by degree, was mistaken. The POE appears now to refer to likelihoods. But likelihoods do not obey the laws of probability, for example, $\text{Lik}(H_{AD}; e) + \text{Lik}(\text{not-}H_{AD}; e)$ need not add to 1 or any number in particular.

Finally, and most emphatically, that the likelihood ratio exceeds 1 does not say a Bayesian posterior probability of H_{AD} exceeds that of not-H_{AD}. That is, statement (i) does not assert

(ii) $P(H_{AD}; e) > P(\text{not-}H_{AD}; e)$.

Yet statement (ii) may sound as if it formally captures the phrase "preponderance of evidence." So there is a good deal of ambiguity here, further showing that legal scholars must clarify, as they have not, the meanings of such evidential standards before we can disentangle arguments about them. For POE to be cashed out as statement (ii) requires a way to assign prior probability assignments (supplemented with an interpretation – frequentist, subjective, conventional). So, the meaning of POE needs settling before we proceed to tackle the question of whether POE and BARD reflect inconsistent SoPs, as Laudan argues. To take some first steps, I continue to draw out conceptual confusions that otherwise hound not just legal discussions, but trials themselves.

1.3 Prosecutor's Fallacies

The intended error-probability construal of BARD, (iii), is not immune to an analogous confusion – most notably, it may be confused with asserting that the probability of the trial outputting an erroneous conviction is .01 (iv). However, the error-probabilistic claim

(iii) $P(\text{trial } T \text{ convicts defendant}; H\text{-defendant is innocent}) = .01$
does not entail a posterior probability statement:
(iv) $P(H\text{-defendant is innocent}; \text{trial } T \text{ convicts defendant}) = .01$.

Statement (iv) requires a prior probability assignment to H-defendant is innocent; statement (iii) does not. Confusing statement (i) with (ii), and (iii) with (iv), appropriately enough, are sometimes known as forms of "prosecutor fallacies"! If philosophers are going to help bring clarity into legal epistemology, they must be careful to avoid classic fallacies, and yet

work in probalistic confirmation theory is often mired in just such confusions (as we saw in Chapter 5). So, we must clean out some of our own closets before we offer ourselves as legal epistemologists.

Whether a given rationale for an SoP holds up depends on how that standard is interpreted. For example, Laudan remarks: "A demanding standard of proof carried by the prosecution, such as proof beyond reasonable doubt, is much more apt to produce false acquittals than false convictions (assuming that defendants are as likely to be innocent as guilty)"(p. 384). Although it is true that if trial T requires BARD (defined as $\alpha \ll \beta$) then it is true that the rate of false acquittals (type II errors) will exceed the rate of false convictions (type I errors) – Laudan's use of "likely" in the parenthetical remark alludes to Bayesian priors (the distribution of guilt among defendants), according to his note 11. The claim, with the Bayesian construal, is that BARD ensures that the proportion of guilty among those acquitted exceeds the proportion of innocents among those convicted:

(v) P(J-defendant is guilty; test T acquits) > P(H-defendant is innocent; test T convicts),

where P is frequentist probability. To grasp statement (v), think of randomly sampling among populations of acquitted and convicted: statement (v) asserts that the probability of the property "guilt" among those acquitted exceeds the probability of the property "innocence" among those convicted. But (v) does not follow from BARD construed as $\alpha \ll \beta$, even assuming the probability of .5 to "innocence" as Laudan does.

It is worth asking how one could justify such a prior probability of .5 to innocence – it may sound like a "fair" assessment, but such is not the case. If it is to be a degree of belief, then this would not do, because we are to presume innocence! Even if one knew the proportion of defendants actually innocent of a crime, this would not give the probability that *this* defendant is guilty (see the exchange with Achinstein, Chapter 5). One might suggest instead using a subjective prior (degree of belief) of 1 to innocence – but then we can get no evidence of guilt. I think any attempt to stipulate a Bayesian interpretation would be ill-suited to capturing the legal standards of proof now used. I assume Laudan agrees, but he seems prepared to relegate such technical matters to one side. What I wish to convince epistemologists of law is that attempting to scrutinize potential conflicts of standards of evidence without also clarifying these probabilistic notions is apt to do more harm than good. Nor is it just the probabilistic notions that are open to equivocal interpretations – the very notion of SoP

is ambiguous, and one of Laudan's most serious allegations turns on this ambiguity, as will be seen in Section 3.

Conversely, there is much to be learned by trying to coherently fix the meanings of these standards of evidence. By striving to do justice to an intended interpretation in the case at hand (affirmative defenses), a position that at first blush seems problematic may turn out, once reinterpreted, to hold water after all. Let us see if this is true in Laudan's case.

2 Interpreting Preponderance of Evidence (POE) in Affirmative Defenses (ADs)

I continue to work on the assumption, following Laudan, that BARD is akin to the error-statistical requirement of controlling the type I error probability to a low value, and with that stipulation, minimizing the type II error probability. What we did not settle is which of the possible construals of POE to use. I will now propose one that I think readily lends itself to the kind of criticisms Laudan wishes to raise regarding SoPs in ADs. Whether his criticism holds is what we need to determine.

Claiming no expertise in the least regarding the legal issues, I will just follow Laudan (acknowledging, as he does, that different states in the United States follow different stipulations). Given that it has been established BARD that the given action was committed by the defendant, the question is whether a given excuse applies, such as self-defense. That is the circumstance under which ADs arise. Questions about whether an AD is warranted may be usefully construed in a manner analogous to questions of whether a test's *intended* low error rates are at least approximately equal to the actual ones. That is, we begin, following Laudan, with the supposition that the error rates of erroneous convictions and erroneous acquittals (α and β) are fixed on societal policy grounds – they are, he says, part of a social contract. We may call these the *primary error rates*; they are the ones chosen for determining the primary question: whether the (presumably illegal) action was committed. By contrast, Laudan's questions about burdens of proof in ADs ask which standards of evidence should apply (for deciding if the act is "excused"). In particular, he asks which AD standards promote or violate the primary error-probabilistic stipulations.

I cannot vouch that this is the thinking behind the legal statutes Laudan criticizes, but I will continue a bit further with this analogy from statistical tests to see where it may lead. In this analogy, substantiating an AD is akin to saving a null hypothesis H_0 from rejection by explaining away the observed anomaly. In the legal setting, H_0 plays the role of an assertion

of "innocence"; analogously, in science, H_0 might be seen as asserting: the observed deviation of the data from some theory T does not discredit T (i.e., "T is innocent of anomalies"). In other words, or so I propose, the AD defense is akin to a *secondary* stage in statistical testing. The secondary stage, recall, concerns testing the underlying assumptions. In particular, it is imagined that a null hypothesis has been rejected with a small p-value – as with the legal case, that much is not in question. A violation of an experimental assumption plays the role of a claim of self-defense because if that excuse is valid, then that invalidates the inference to guilt. (Similarly, if the legal excuse is valid, the defendant is legally "innocent" – therefore not guilty – even though the action is not contested.)

Laudan wants us to focus on the rationale for requiring the defendant to prove a POE that his proposed defense or excuse is true. I focus on self-defense. Laudan's worry is that requiring a defendant to provide a POE for the excuse of self-defense seems to vitiate the requirement that the prosecutor provide evidence BARD of guilt. Now Laudan focuses on the small type I error rate that accompanies BARD – rate of erroneous conviction – but to tackle our present concern we must consider the rate of type II errors. Although the rate of a type I error (erroneous conviction) is set by the "social contract" to be smaller than that of the type II error (erroneous acquittals), surely we would not abide by large type II error rates, or even a type II error rate as large as .51. That would mean failing to convict a guilty person more than 50% of the time. Once it is remembered that the type II error rate must also be sensibly low, even if several times greater than the type I error rate, the arguments about AD can be more meaningfully raised (e.g., a type II error rate ten times that of the type I error rate associated with BARD would be approximately .1 or .2). If there is too much latitude in permitting such excuses, then the originally intended low type II error rate does not hold – the actual type II error rate can become too large.

Continuing to focus on the AD of self-defense, an overly promiscuous standard for excusing would clearly lead to a high (primary) type II error rate – the question is what is overly promiscuous. The answer would seem pretty plainly to be any standard that invalidates the primary requirement to avoid a high type II error rate. We can well imagine that, were there blatant grounds for suspecting self-defense, the case would scarcely have been brought to court to begin with (e.g., shooting a student who was gunning down people in a classroom).

Because we may assume that in realistic ADs the explanation for the "anomalous" data or action is not so obvious, it stands to reason that some

evidence is needed. Were it sufficient for the defendant to give an excuse that is always available to a defendant whether guilty or innocent (e.g., a variation on "I believed I had to defend myself") – much as Velikovsky can always save himself from anomaly by appealing to amnesia (see Chapter 4) – the type II error rate would be considerably raised. I am not saying whether a given handling of AD would permit the type II error rate to increase, but rather I suggest that this would be a productive way to address the concern Laudan raises.

A good analogy is the reasoning about the cause of the observed deflection effect – anomalous for Newton's law. "The many Newtonian defenders adduced any number of factors to explain the eclipse effect so as to save Newton's law of gravity" (Mayo, 1996, p. 287). Suppose all they had to do was assert that the observed deflection is due to some Newton-saving factor N (e.g., shadow effect, corona effect, etc.), without giving positive evidence to corroborate it. The result would be a very high probability of erroneously saving Newton. How severely the proposed excuse would need to pass to avoid overly high type II error rates, even approximately and qualitatively determined, is what would matter.

3 Is the Interpretation of Evidence Relative to Costs of Errors?

In analyzing standards of evidence in law, Laudan elicits a number of novel insights for epistemology and philosophy of science. The variable standards of proof in the law, he observes, stand in contrast to the classical image of evidential appraisal in science. Laudan draws some dire lessons for the practice of science from variable standards of proof in the law that I want now to consider:

If there are generally no canonical rules for the acceptance of scientific theories, if standards of acceptance vary from one scientist to another within a given specialty and from one science to another (as they patently do), then how do we go about describing and defending the conviction that science is a quintessentially rational, rule-governed activity? (Laudan, this volume, p. 395)

I would concur in denying that scientists apply uniform rules for the acceptance of scientific theories, but I find it no more plausible to suppose that scientists possess individual rules for accepting theories on the basis of their chosen cost-benefit analysis. I do not see scientists going around applying rules for theory acceptance in the first place – whether theory acceptance is given a realist or an antirealist interpretation. Given that Laudan's discussion here is about standards of evidence for reaching an inference

(e.g., grounds to acquit, grounds to convict), I take him to be saying roughly the following:

Laudan's variable standards thesis: The variability of standards of evidence, growing out of the differing assessments of costs of mistaken inferences, poses a serious threat to the conviction that there is a single standard for determining what conclusions are warranted from evidence.

Considering Laudan's variable standards thesis presents us with the opportunity to consider the possibility of objective criteria for warranted evidence and inference more generally. Laudan's provocative charge seems to be that there are no uniform, overriding standards for scrutinizing the evidential warrant of hypotheses and claims of interest; any such standards are determined by choices of errors and the trade-offs between costs of errors. Note that his charge, if it is to carry a genuine provocation, must not be, that individual scientists prefer different methods or worry more or less about different errors, much less that there invariably are different personal benefits or harms that might accrue from a hypothesis being warranted by evidence. These facts do not compel the conclusion that I take Laudan to end up with: that evidential warrant is relative to cost-benefit assignments varying across scientists and fields. We may call this view that of the "cost relativism" of standards of evidence.

But do standards of evidence patently vary across sciences? The only way Laudan's provocative thesis would follow is by trading on an equivocation in the meaning of "standard of evidence" – one that is perhaps encouraged by the use of this term in the legal context. If a "rational rule" or a "standard of evidence" is understood to encompass costs and utilities of various sorts, then, given utilities vary, rational rules vary. For example, we know that in the U.S., OSHA operates with different standards than does the EPA: for the former workplace settings, a statute may set the risk increase before a suspected toxin is open to a given regulation as 1 additional cancer per 10,000 – anything worse is declared an "unacceptable risk." The EPA operates with a more stringent standard, say 1 additional cancer in 1,000. The *agency standards* operate with different cut-offs for "unacceptable risks." (Such discretionary judgments are sometimes referred to as risk assessment policy options; see Mayo and Hollander, 1991.) But this does not mean the two agencies operate with different "standards of evidence" understood as criteria for determining whether the data warrant inferring a .0001 risk or a .001 risk or some other risk. If they did, it would be impossible to discern with any objectivity whether their tests are sensitive enough to inform us whether the agency's standards (for acceptable risk) are being met!

That is why Rachelle Hollander and I (1991) explicitly distinguish "acceptable risk," which involves a policy or value-laden judgment about costs, from "acceptable *evidence* of risk," which does not. Given the chosen agency standard, whether the data do or do not indicate that it is met is a matter of the extent of the evidence. The equivocation is exacerbated in the legal setting because there the term "standard of proof" is used to refer to a social policy judgment. Debating whether to require BARD or POE for a given legal context is to debate a policy question about how high to set the bar. To emphasize this legal usage of "standard of proof," we write it as SoP. A distinct evidential standard would refer to the standards of evidence for determining if a given policy standard is met. Once the SoP is chosen, whether or not given evidence reaches the standard chosen is not itself a legal policy question. As difficult as it may be to answer it, the question is a matter of whether the evidence meets the standard of stringency given by the error probability corresponding to the SoP chosen on policy grounds.

The analogy with risk assessment policy options is instructive. In legal epistemology an SoP is precisely analogous to a standard of acceptable risk – it is understood as a risk management concept. Altering a level of acceptable risk, we may grant, alters what counts as a "rational" policy or decision (e.g., reduce exposure to the toxin in the case of an EPA statute; perhaps "do nothing" in the case of an OSHA statute). But then Laudan's claim about the existence of different risk management settings becomes trivially true and would scarcely rise to the height of threatening "the edicts of science as the embodiment of rationality," as Laudan declares. Setting the SoP in cases of evidence-based policy is a matter of social, economic, pragmatic, and ethical values, but, given those specifications (e.g., given by fixing error rates), whether an inference is warranted is not itself relative to those values.

Without deciding whether Laudan really means to be espousing "cost relativism," we can pursue my argument by considering an analogous position that frequently arises in the area of science-based policy. It is often couched in terms of claims that there are different types of rationality, in particular "scientific rationality" and something more like "ethical rationality." The argument comes in the form of statistical significance tests, and they link up immediately with previous discussions.

In science, the null hypothesis is often associated with the status quo or currently well-supported hypothesis, whereas a challenger theory is the rival alternative.

(Scientific context) H_0: No anomalies exist for an established theory *T*.

The small type I error rate serves to make it difficult for the tried-and-true accepted hypothesis to be too readily rejected. In the realm of risk assessment, some argue that such standards, while fine for "scientific rationality," may be at odds with "ethical rationality." In particular, suppose our null hypothesis is the following.

(Risk policy context) H_0**:** No increased risks are associated with drug X.

Because the consequences of erroneous acceptance of H_0 would lead to serious harms, it may be recommended that, in risk policy contexts, it is the type II error probability that needs to be controlled to a small value, since this approach would be more protective.[2]

The concern is that, by securing so small a probability of erroneously declaring an innocent drug guilty of increasing a risk (type I error), studies may have too high a probability of retaining H_0 (the drug is innocent) even if a risk increase of δ is present:

The probability of a type II error: $\beta(\delta) = P(\text{test } T \text{ accepts } H_0; \text{ increased risk } \delta \text{ is present})$.

The concern is with cases where the probability of a type II error is high. We should use this information to argue via the weak severity principle:

If P(test T accepts H_0; increased risk δ is present) is very high, then accepting H_0 with test T is poor evidence that an increased risk δ is absent.

Although H_0 "passed" test T, the test it passed was not severe – it is very probable that H_0 would pass this test even if the increased risk is actually as large as δ (see Chapter 7).

The ability to critique which risks are or are not warranted is the basis for avoiding "cost relativism." Laudan's own critique of the consequences of varying legal standards would seem to require this.

4 Recommendations for Legal Epistemology

Laudan's contribution provides valuable grist for three issues that would arise in an epistemology of law. First is the need to clarify the standards of evidence. In a formally specified problem, there are relationships between equally probable hypotheses, equally likely hypotheses, and equal error probabilities (in the sense of frequentist statistics), but they say very different

---

[2] In fact this should be the nonnull hypothesis, because erroneously inferring the drug is safe would be deemed more "serious," but in practice it is typically the null hypothesis.

things and, unless these notions are pinned down, it will be difficult to evaluate arguments about what consequences for error rates would accrue from adopting one or another standard of proof. Second, there is a need to address Laudan's questions about the consistency/inconsistency of legal SoPs, in particular, standards for ADs and the intended BARD standard for guilt and innocence. I have proposed we do so by considering whether they ensure or violate the stipulated error rates for guilt/innocence. Third, there is a need to distinguish the specification of acceptable risks of error – an issue of policy or management – from that of appraising the acceptability of the evidence, given those standards.

References

Laudan, L. (2006), *Truth, Error and Criminal Law: An Essay in Legal Epistemology*, Cambridge University Press, Cambridge.

Mayo, D.G. (1996), *Error and the Growth of Experimental Knowledge*, University of Chicago Press, Chicago.

Mayo, D.G., and Hollander, R., (eds.) (1991), *Acceptable Evidence: Science and Values in Risk Management*, Oxford University Press, New York.