Synthese (2008) 163:305–314
DOI 10.1007/s11229-007-9303-5

The error statistical philosopher as normative naturalist

Deborah Mayo · Jean Miller

Received: 9 November 2007 / Accepted: 9 November 2007 / Published online: 11 March 2008

Abstract We argue for a naturalistic account for appraising scientific methods that carries non-trivial normative force. We develop our approach by comparison with Laudan's (American Philosophical Quarterly 24:19–31, 1987, Philosophy of Science 57:20–33, 1990) "normative naturalism" based on correlating means (various scientific methods) with ends (e.g., reliability). We argue that such a meta-methodology based on means–ends correlations is unreliable and cannot achieve its normative goals. We suggest another approach for meta-methodology based on a conglomeration of tools and strategies (from statistical modeling, experimental design, and related fields) that affords forward looking procedures for learning from error and for controlling error. The resulting "error statistical" appraisal is empirical—methods are appraised by examining their capacities to control error. At the same time, this account is normative, in that the strategies that pass muster are claims about how actually to proceed in given contexts to reach reliable inferences from limited data.

Keywords Normative naturalism · Reliable inference · Methodology · Meta-methodology · Error statistics · Learning from error · Controlling error · Appraising scientific methods · Laudan · Canonical models of error · Means–ends approaches

D. Mayo
Department of Philosophy, Virginia Tech, 235 Major Williams Hall (0126),
Blacksburg, VA 24061, USA
e-mail: mayod@vt.edu

J. Miller (✉)
Department of Philosophy, Virginia Tech, 211 Major Williams Hall (0126),
Blacksburg, VA 24061, USA
e-mail: jemille6@vt.edu

1 Introduction

In reaction to the shortcomings of logical empiricist prescriptive accounts of science, many philosophers of science have tended to view the task of philosophy of science as largely, if not solely, descriptive and not normative.[1] These philosophers of science hope that by appropriately attending to the details of actual science, whether through historical episodes or contemporary practice, it may be possible to extract philosophical lessons about scientific knowledge. Looking to historical practice, however, reveals norms that are so various, and changing that many have come to deny any general claims can be made about how scientists acquire knowledge save perhaps, as Alan Chalmers suggests, trivial platitudes such as "take evidence seriously (1999, p. 171 but see Mayo 2000 for a rejoinder)." At the same time, many philosophers of science (rightly) feel that to give up the normative goal entirely is to forfeit the critical stance philosophers would like to sustain, and invites all manner of relativisms and social constructivisms (e.g., Achinstein, Chalmers, Giere, Hacking, Kitcher, Laudan, Worrall).

Is it possible to have an account of the methodology of science that is empirical and also genuinely (and non-trivially) normative, and if so, how? How can philosophers of science, usefully, be normative naturalists? Doubtless most, if not all, of the philosophers of science listed above think they *can* extract normative lessons from history or one or more sciences—be it biology, anthropology, economics, psychology or cognitive science. However, their projects depend on one or another assumptions (e.g., about what counts as science or rational, about the success of science, or about self-correcting processes of cognition) which are either accepted without empirical grounding, or else give rise to charges of circularity. Even these fundamental concerns, however, would diminish in importance if these approaches delivered the goods, but thus far they have not. The entire "history and philosophy of science" movement, for example—although providing fascinating historical details—has failed to deliver on its promise of explaining what makes science so successful, much less providing tools to improve our strategies for how to obtain, and learn from, reliable evidence.

Nevertheless, we think we should take the normative naturalist lessons seriously. We should look to the practice of science, but not merely to some famously neat historical episodes which handily "fit" a variety of accounts of confirmation and testing. Rather, we should deliberately examine cases where data are inexact, noisy and incomplete, and where a host of errors and biases loom—the very contexts that belie the tidy accounts proffered by the logical empiricists. When we do, we see scientists appeal to a conglomeration of research methods for collecting, modeling, and learning from data in the face of limitations, uncertainties, and the threats of errors—including modeling techniques, probabilistic and computer methods—all of which we can house under a very general rubric of the *methodology of statistical modeling and inference.* So this would seem a natural, yet still untapped, place for normative naturalists to begin.

We will be directed not to identifying particular recommendations the scientist seeks for a specific inquiry (e.g., sample size, sampling method) but to achieving the

[1] Thomas Kuhn's work was a major influence.

more general goals of the normative naturalist: to understand and improve our strategies and methods for making reliable inferences in the kinds of contexts and inquiries we actually confront. "Naturalists," Kitcher remarks, "view members of our species as highly fallible cognitive systems (1992, p. 58)." But to extract the philosophical lessons he and others are after, we think, requires going beyond his call for studying human psychological and biological capacities and limitations, to studying the most sophisticated means our species has devised for obtaining knowledge in spite of those limitations. Although there is not just one unified science to encompass the strategies and methods we have in mind, one may trace out (from statistics, experimental design, empirical modeling, computer simulation, etc.) a methodology of interconnected methods that can be described, we claim, as methods for *learning by controlling error*. To have a ready way to refer to the conglomeration of tools and strategies for learning from error we may call it *error statistics.*

In this discussion, we (1) begin by considering Laudan's normative naturalism, (2) compare it to the recommendations of the error statistical meta-methodology, and (3) examine some questions and criticisms to which the error statistician must respond.

2 Laudan's (1987, 1990) normative naturalism as assessing means–ends covariances

We share two fundamental theses that Laudan argues underlie all normative naturalisms:

- Normative Aim: "[T]he chief aim of the methodological enterprise is to discover the most effective strategies for investigating the natural world."
- Naturalistic Methodology: "Rules for learning from experiment are empirical and may be evaluated empirically" (Laudan 1987, p. 27).

Despite agreeing with these two goals, we believe there are several impediments to achieving them by means of Laudan's approach, and other analogous approaches. According to Laudan:

- normative rules of epistemology are best construed as *hypothetical imperatives*, linking means and ends;
- the soundness of such prudential imperatives depends on certain *empirical* claims about the connections between means and ends;
- accordingly, empirical information about the relative frequencies with which various epistemic means are likely to promote sundry epistemic ends is a crucial desideratum for deciding on the correctness of epistemic rules (1990, p. 46).

Some examples of methods that Laudan gives are:

(i) propound only falsifiable theories
(ii) avoid ad hoc modifications
(iii) prefer theories which make successful surprising predictions over theories which explain only what is already known.
(iv) when experimenting on human subjects, use blinded experimental techniques.
(v) use controlled experiments for testing causal hypotheses (Laudan 1987, p. 23).

The "means/ends" formulation is obtained by viewing methods such as (i)–(v) as the means for obtaining successful or reliable theories. For example, methods (ii) and (iii) become:

(ii) if you want theories with high predictive reliability, reject ad hoc hypotheses
(iii) if you want theories likely to stand up successfully to subsequent testing, then accept only theories which have successfully made surprising predictions.

However, testing a methodological rule already requires making use of some rule of appraisal, thereby opening Laudan's "meta-methodological" project to the charge of circularity. Nevertheless, Laudan argues that since all theories of methodology share something like arguing from instances or enumerative induction, we can use a form of that rule—at least to get the meta-methodological project off the ground. He formulates such a rule as R_1:

> (R_1): If actions of a particular sort, M, have consistently promoted certain cognitive ends, e, in the past, and rival actions, N, have failed to do so, then assume that future actions following the rule 'if your aim is e, you ought to do M' are more likely to promote those ends than actions based on the rule 'if your aim is e, you ought to do N' (Laudan 1987, p. 25, m and n replaced with M and N).

To determine whether one methodology is better than the other, Laudan claims, "we simply inquire about which methods have promoted, or failed to promote, which sorts of cognitive ends in the past" (1987, p. 27).

What kind of data can ensure that the antecedent of R_1 is warranted? Since, as Laudan recognizes, the methodological rules are statistical, evidence for the antecedent consists of finding that method M promotes end *e* more often than rival methods *in some sample of cases*. But how do we know we have found even a single instance where applying M promoted reliability? Putting aside the problem of ascertaining if a given method M has actually even been applied, we are left with the more serious problem of determining whether it was the application of method M that promoted the end of reliability. More generally, finding an instance in which method M is satisfied (violated) and cognitive end *e* reached (not reached) fails to tell us that reaching (failing to reach) the end was due to satisfying (violating) the rule.[2]

The problem of "false negatives" looms especially large. Using R_1, Laudan claims to have "discredited" rules (ii) and (iii) (on the value of avoiding ad hoc hypotheses and requiring surprising predictions (1987, p. 26)). However, the available historical record is a highly unrepresentative sample of applications of methodological rules with regard to the question of reliability. Thus, the extent to which violating a given rule creates an obstacle to achieving the goal of reliability—the very thing we would be looking for—would be missed. There would be scant evidence of cases where violating a rule resulted in an unreliable theory because, by and large, if the theory were not wiped off the books altogether, it would have been shored up by other means.

[2] To put this in other words, the problem is not with inferring that methods that worked in the past will work in future; the problem is using a sample of (means/end) correlations to determine if the methods worked in the first place. Other discussions of Laudan's normative naturalism include Kitcher (1992), Leplin (1990), Mayo (1999), Rosenberg (1990), Worrall (1999).

Note too that such negative evidence fails to provide precisely the kind of normative advice that Laudan rightly sees as the job of a normative meta-methodology. Being told, for example, that if your aim is reliability, you need not (or should not?) use method M is hardly to give positive information about what effective strategies you *should* use.

In sum, without knowing whether satisfying a rule is the, or even a contributory, cause of the goodness of a resulting theory, and conversely, whether violating the rule is the reason that a theory fails to be reliable, information about correlations, even where obtainable, does not help. A given theory may be seen as resulting from applying any number of methods. Laudan's meta-methodology could achieve what he wants it to only if it were supplemented with a way to determine whether the method was causally efficacious in the context in question. This demands a separate method for determining the capacities of the methods. But if one had such a separate method for determining whether and how rules work to achieve ends, then one would not need to collect correlational evidence in the first place. Any naturalistic method that strives for normativity will similarly require getting into the "black box" of experimental methods, so let us turn to an approach that sets the stage for accomplishing this.

3 The error statistical (ES) approach to meta-methodology

What we are calling the ES approach[3] takes seriously the fundamental posit of naturalistic epistemology: that methodological claims, being empirical, should be appraised in the way we appraise claims in science and common sense. In fact, a good deal is known about how to evaluate means–end strategies, and we know that we would not settle for correlational data if we could actually uncover the causal mechanisms. Uncovering "the mechanism" in the case of methodological rules requires uncovering the properties inherent in a method that enables its use to detect errors, and either subtract them out or control them to achieve reliable inferences. Although the efficacy of these methods are *context specific*, there is a general goal that they serve to promote; namely, to provide reliable procedures for arriving at claims.[4]

3.1 Severity as a principle of evidence

Let us propose the following principle of evidence:

> One has satisfactory evidence for a hypothesis or claim only to the extent that the ways the claim could be incorrect have been well probed by procedures that almost surely (i.e., with very high probability) would have signaled the presence of the errors, if they existed, and yet no errors are found; instead, our observations

3 See Mayo and Spanos forthcoming for a full discussion of error statistics in both statistics and philosophy of science.

4 Some epistemologists might prefer to describe them as procedures for arriving at beliefs. Doing so links the project with reliabilists, e.g., Dreske, Goldman, Kornblith.

are of the sort that are to be expected given the *absence* of the error.[5] In such cases we can say that the error has been reliably or *severely probed.*

For example, if a student has passed numerous, probative, logic exams, we would say we are warranted in inferring she knows at least most of the material; we reason that if this were not so, she almost surely would not have scored so well on all of these varied logic exams.

The ES methodologist would appraise methods according to how well they afford *severe tests* and error probes (see for example, Bartz-Beielstein, Parker, and Taper et al., herein). Different inferences will demand that different errors be severely probed before regarding data as providing evidence for them—so we can explain why no single method will do. Nevertheless, from the ES perspective, the overarching aim *of any* inference—including inferences to claims about data themselves—is the ability to show that the relevant errors have been satisfactorily probed, and either found absent, subtracted out, or accounted for in the final inference.[6] Methodological rules, therefore, are appraised according to how well they promote these aims.

3.2 Models of error

To illustrate, consider how we would test the value of experimental control of relevant factors in reaching a causal inference (e.g., that treatment T causes an effect E). One can reason out why literal control of all factors would be desirable for this causal inference. If we can neutralize all factors except the one being tested, T, then we can reliably argue from E's occurrence that T causes E. Although full control of all possible confounding factors is rarely if ever possible, it is an idea that provides a *model* or *exemplar* for pinpointing the rationale for control. Reference to this model guides scientists to find procedures to approximate control, so as to allow arguing *as if* they had full control, (e.g., randomized control groups). Following the procedure of randomized control groups allows experimenters to argue although other factors are operating, they affect both groups equally. Thus, the effects of confounding factors, whatever they are, are cancelled out.

Note that in legitimizing relatively uncontroversial rules about the effectiveness of experimental methods (e.g., of controls to avoid confounding factors), there is no appeal to how often they have led to correct causal hypotheses as Laudan's means–ends tests would seem to require. *Instead, there is an understanding of how they function in avoiding key impediments* to sustaining the inference, in accordance with the ES principle of evidence above. Granted, questions about how methods function, and whether they will work in a given case, are empirical: they concern real problems in actual empirical inquiries. Nevertheless, their answers are not arrived at by checking how often methods correlate with a reliable end-product. The same goes for methodological rules in general. We agree with Clark Glymour (2008) who writes: "What methodolo-

[5] One may want to distinguish the ability of a person to show the error is ruled out, from the error having been ruled out—whether or not the person can show it. We leave this to one side here.

[6] See Mayo (2002, 2008a) for a discussion on extending severity to high level theory appraisal, also Staley herein.

gists ought to provide, if they can, is effective methods justified by demonstrations of their reliability, if not in every possible circumstance, then in circumstances as general as can be found." The error statistical approach guides methodologists on how to accomplish this.

We reasoned out the efficacy of experimental controls in an entirely informal manner; other cases call for appealing to formal statistical arguments. In polling methods, for example, we want to infer from a random sample of U.S. citizens the proportion in the full population who favor a given policy (e.g., going to war with Iraq). The procedure of forming an interval estimate with a 1% margin of error *has the property* of outputting correct estimates 99% of the time. It is a property of the rule or procedure, assuming the statistical assumptions (e.g., random sampling) are satisfied. The idea that we may look to the (objective) properties of methods is the cornerstone of the ES approach to methodology. As such, statistical ideas are relevant for our meta-methodological task because of their ability to teach us about the error probing properties of tests and other inferential procedures.[7]

4 Can we avoid the charges against traditional normative naturalisms?

According to Worrall, any fully naturalized philosophy of science, any account that endorses the complete rejection of the a priori "either runs into logical circles (and hence incoherence) or ends up in a particularly unpalatable form of relativism (Worrall 1999, p. 339)." The basis for these charges is that a fully naturalized philosophy of science needs to make use of empirical methods in deciding what sciences or scientists to study, and in deciding how to extract epistemologically relevant lessons. If it postulates one method (or meta-methodology) as superior, it would seem to be assuming what it was supposed to provide. Otherwise it would seem to be conceding that one's approach is relative to a particular choice of (fallible) empirical theory or epistemological principle (e.g., study biology and not witchcraft).[8]

The normative naturalist, Worrall concludes must allow for what he calls "minimal *a priorism*;" and his own choice of a priori starting point consists basically of Laudan's rule R_1, and the preference for novel predictions (rule of novelty). In our view, no matter how one obtains such rules, they too must be scrutinized (e.g., Mayo (2006) on Critical Rationalism). We have already argued against taking R_1 as a reliable rule to apply for appraising methodological rules. Elsewhere, the (highly contested) principle

[7] Although strictly relevant only to tests modeled as random variables, statistical error assessments also serve as exemplars for qualitative assessments. They usefully illuminate (see Mayo 1996) the ways in which a variety of specific errors may all be grouped around a handful of error types: errors about a causal factor, an experimental artifact or chance effect, a quantity or value of a parameter, and errors about assumptions of the experimental data (e.g., claiming a sample is random when it is not, thinking the microscope is working when it is not).

[8] Empirical studies by psychologists (and economists) to try to understand how people reason are open to analogous challenges. We would wish to scrutinize their most basic assumptions, if they are to be used for the normative project: What is the basis for assuming that the various studies on subjects afford adequate data to arrive at warranted hypotheses about ways people ought to reason? What are the assumptions underlying those studies? About probabilistic inference? About the "right" answer? About how to extrapolate to improved performance in science? See Mayo (2008b).

of novel prediction is shown to be neither necessary nor sufficient for severe or reliable tests.[9]

In appealing to statistical methodology for procedures to generate, model, and obtain reliable inferences and severe error probes, we are appealing to methods that are effective for the learning contexts *of this* world: it is an empirical not an a priori matter. On the one hand, the statistical properties are deducible mathematically, e.g., a statistically significant result at the 0.05 level would occur no more than 5% of the time, if the null hypothesis is true entails 95% of the time that the test will correctly signal that it is a mistake to rule out the "chance" explanation for an observed difference.[10] On the other hand, in order to apply a statistical test to an actual inquiry the specific data must satisfy, at least approximately, the formal statistical assumptions of the model. Although it follows that usable data for statistical inference will themselves have empirical assumptions, a regress is avoided. We can arrive at more accurate data from highly shaky measurements (e.g., by averaging and other statistical techniques). By deliberately varying assumptions of model validation techniques, one can fortify and self-correct data ensuring reliable statistical inferences from individually unreliable and inaccurate data (see Mayo and Spanos 2004).

But how do we justify the ES meta-methodology and principle of evidence? We began with the normative naturalist's goal of arriving at improved procedures for learning despite fallibility, error and limited data. An adequate account needs to provide methods for how to proceed when confronted with a learning problem, not merely tools to reconstruct, after-the-fact, an episode so that it may be deemed rational. These goals, we think, underwrite the error statistical approach to normative naturalism, and its recommendations may be appraised accordingly. While it may be regarded as more of a program than a finished approach, we hope to encourage further exploration of some of the untapped ES ideas for moving toward an adequate normative naturalistic philosophy of science.

References

Achinstein, P. (1998). Why philosophical theories of evidence Are (and Ought to Be) ignored by scientists. *Philosophy of Science, 67* (Proceedings), S180–S192.

Achinstein, P. (2008). Mills' sins. In D. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on the philosophy of science, inductive-statistical inference, and reliable evidence*. Cambridge: Cambridge University Press (forthcoming).

Bartz-Beilestein, T. (2008). How experimental algorthmics can benefits from Mayo's extensions to Neymoan-Pearson theory of testing (this volume).

Chalmers, A. (1999). *What is this thing called science*? (3rd ed.). Indianapolis/Cambridge: Hackett Publishing Company, Inc.

9 See Mayo (1991, 1996). One can contrast her ES critique of the novelty rule with the "instance disconfirmation" in Laudan's use of R_1 discussed above. Confusions about novelty and double counting are addressed at length in Mayo forthcoming. Worrall's way of dealing with these counterexamples to the necessity of use-novelty has undergone various changes (2002, 2008).

10 That is, 95% of the time the test will correctly accept the null hypothesis asserting the difference is "due to chance."

Chalmers, A. (2008). In D. Mayo, & A. Spanos (Eds.), *Error and inference: Recent exchanges on the philosophy of science, inductive-statistical inference, and reliable evidence*. Cambridge: Cambridge University Press (forthcoming).
Dreske, F. (1985). Precis of knowledge and the flow of information. In H. Kornblith (Ed.), *Naturalizing epistemology* (pp. 169–188). Cambridge, Mass.: MIT Press.
Dreske, F. (2000). *Perception. Knowledge and belief: Selected essays*. Cambridge: Cambridge University Press.
Giere, R. (1996). Scientific inference: Two points of view, *Philosophy of Science, 64* (Proceedings), S180–S184.
Giere, R. (2004). Critical hypothetical evolutionary naturalism. In C. Heyes & D. L. Hull (Eds.), *Selection theory and social construction: The evolutionary naturalistic epistemology of Donald T. Campbell* (pp. 53–70). Albany: State University of New York.
Giere, R. (2006). *Scientific perspectivism*. Chicago: University of Chicago Press.
Glymour, C. (2008). Searching for explanations, searching for truths. In D. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on the philosophy of science, inductive-statistical inference, and reliable evidence*. Cambridge: Cambridge University Press (forthcoming).
Goldman, A. (1985). What is justified belief? In H. Kornblith (Ed.), *Naturalizing epistemology* (pp. 91–114). Cambridge, Mass.: MIT Press.
Goldman, A. (2002). *Pathways to knowledge: Private and public.* New York: Oxford University Press.
Kitcher, P. (1992). The naturalists return. *The Philosophical Review, 101*(1), 53–114.
Kornblith, H. (1985). Introduction: What is naturalistic epistemology? In H. Kornblith (Ed.), *Naturalizing epistemology* (pp. 1–14). Cambridge, Mass.: MIT Press.
Kornblith, H. (2002). *Knowledge and its place in nature*. Oxford: Oxford University Press; New York: Clarendon Press.
Kuhn, T. (1962, 1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
Laudan, L. (1987). Progress or rationality? The prospects for normative naturalism. *American Philosophical Quarterly, 24*, 19–31.
Laudan, L. (1990). Normative naturalism. *Philosophy of Science, 57*, 44–59.
Leplin, J. (1990). Renormalizing epistemology. *Philosophy of Science, 57*, 20–33.
Mayo, D. (1991). Novel evidence and severe tests. *Philosophy of Science, 58*, 523–552. (Reprinted in *The Philosopher's Annual, XIV*, 203–232).
Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago: The University of Chicago Press.
Mayo, D. (1999). Making progress with Laudan's Problems: 1977–1997. In *Progress and its problems after 20 years*. The National Autonomous University of Mexico (in Spanish).
Mayo, D. (2000). What is this thing called philosophy of science? In Review symposium of A. Chalmers' What is this thing called Science? *Metascience, 9* (July 2000), 179–188.
Mayo, D. (2002). Theory testing, statistical methodology, and the growth of experimental knowledge. In *The scope of logic. Methodology and philosophy of science*: *Volume One of the 11th International Congress of Logic, Methodology and Philosophy of Science, Cracow, August 1999*. Dordrecht: Kluwer Academic Publishers.
Mayo, D. (2006). Critical rationalism and its failure to withstand critical scrutiny. In C. Cheyne & J. Worrall (Eds.), *Rationality and reality: Conversations with Alan Musgrave*, Kluwer sereis studies in the histroy and philosophy of science (pp. 63–99). The Netherlands: Springer.
Mayo, D. (2008a). Severe testing, error statistics, and the growth of theoretical knowledge. In D. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on the Philosophy of Science, Inductive-Statistical Inference, and Reliable Evidence*. Cambridge: Cambridge University Press (forthcoming).
Mayo, D. (2008b). External validity and the rational scrutiny of models of rationality. *Philosophy of Science* (Proceedings), forthcoming.
Mayo, D. (forthcoming). How to discount double counting when it counts. *The British Journal for the Philosophy of Science.*
Mayo, D., & Spanos, A. (2004). Methodology in practice: statistical misspecification testing. *Philosophy of Science(Symposia), 71*, 1007–1025.
Mayo, D., & Spanos, A. (Eds.) (2008). *Error and inference: Recent exchanges on the philosophy of science, inductive-statistical inference, and reliable evidence*. Cambridge: Cambridge University Press (forthcoming).

Mayo, D., & Spanos, A. (forthcoming). Error statistics. In P. S. Bandyopadhyay & M. Forster (Eds.), *The handbook of philosophy of science* (7). Elsevier: Amsterdam, The Netherlands.
Parker, W. (2008). Computer simulation through an error-statistical lens.
Rosenberg, A. (1990). Normative naturalism and the role of philosophy. *Philosophy of Science, 57*, 34–43.
Staley, K. (2008). Error-statistical elimination of alternative hypotheses (This volume).
Taper, M. Staples, D. F., & Shepard, B. B. (2008). Model structure adequacy analysis: Selecting models on the basis of their ability to answer scientific questions (this volume).
Worrall, J. (1999). Two cheers for naturalized philosophy of science—or: Why naturalized philosophy of science is not the cat's whiskers. *Science & Education, 8*(4), 339–361.
Worrall, J. (2002). New evidence for old. In P. Gardenfors, J. Wolenski, & K. Kijania-Placek (Eds.), *The scope of logic. Methodology and philosophy of science: Volume One of the* 11th *International Congress of Logic, Methodology and Philosophy of Science, Cracow, August 1999.* Dordrecht: Kluwer Academic Publishers.
Worrall, J. (2008). Error, tests and theory confirmation. In D. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on the philosophy of science, inductive-statistical inference, and reliable evidence*. Cambridge: Cambridge University Press (forthcoming).