Suppe, ed. Urbana: University of Illinois Press, pp. 266-83.

Suppes, Patrick (1974b), "Probabilistic Metaphysics," *Filosofiska Studier*, 22. Uppsala: Uppsala Universitet. □

Michael Gardner/Barry Loewer, Philosophy, University of South Carolina, Columbia, South Carolina 29208, USA

Understanding *Understanding Scientific Reasoning*,

Response to Gardner and Loewer

RONALD N. GIERE

I appreciate the opportunity to reply to Gardner and Loewer's review. It provides me a chance to acknowledge in print the comments they have sent me over the past several years. It also provides a chance to respond indirectly to a number of others who have written me with similar questions but have not yet received an individual reply.

Writing a textbook requires balancing a number of often conflicting demands. One is to be true to the subject matter. This is especially difficult in philosophy since on many topics there is little consensus on what are the correct views. Another demand is that the material be presented in such a way as to be accessible to the intended audience—in this case undergraduates, particularly lower division students who may not yet have taken a college level course in either philosophy or science. In addition, one must produce a book that a reasonable number of teachers will want to use. This is connected with the demands of editors and publishers for a book with a reasonable market. Finally there are the demands of time and energy which determine when one decides that enough is enough.

In writing *Understanding Scientific Reasoning* (USR) I became painfully aware of all these demands. In general I tried to give priority to the potential *student*. I wanted a book that would provide students with reasoning skills that they could apply in comprehending and evaluating scientific information of the type one meets in everyday life. In short, I wanted a book that would do for scientific reasoning what some logic texts attempt to do for reasoning in general. In pursuing this aim I deliberately suppressed the desire to get students to think *philosophically*—to get them to do philosophy of science. It is enough, I thought, if one can get them to think "scientifically." Elementary logic courses do not generally take up the philosophy of logic. Why must a course in scientific reasoning take up the philosophy of science? Of course there is an implicit philosophy of logic behind every logic text. And USR obviously contains an implicit philosophy of science. But the philosophy of science is intended primarily as a means to the end of teaching reasoning skills. It is not presented as the subject for study. This orientation is perhaps most evident in the exercises, few of which focus on the philosophical framework. Most require students to *use* the framework in analyzing a report of some scientific study. My exams follow the same format.

On the other hand, USR can be—and has been—used quite successfully as a text for a course in the philosophy of science. This requires that the instructor use the framework of the text as a foil for other possible views of theories, causality, evidence, etc. And this in turn requires that the instructor be prepared to present such other views. But one must be clear at the outset what course one is teaching—and plan accordingly. That is the main thing I have learned in talking to people who have used USR in their own courses. Those who were uncertain of whether they were teaching reasoning skills or the philosophy of science ended up doing some of both, but were often unhappy with the results. This is understandable. Even with a full fifteen weeks there is not enough time to get through the text if one stops to examine alternative philosophical frameworks. Unless one plans very carefully, there is considerable danger of ending the course with the students somewhere in the middle of the chapter on probability or the testing of statistical hypotheses. If this happens, the students will have struggled with some new and occasionally difficult concepts without ever getting to apply these ideas to the inherently much more interesting and practically important *causal* hypotheses treated in Chapter 12. Thus, failing to be rewarded for their efforts, the students end up unhappy with the course.

It is clear that Gardner and Loewer are mainly concerned with the *philosophical* adequacy of the framework developed in USR. At times their review reads more like a referee's report on a philosophical paper than a review of an elementary text for an audience concerned primarily with teaching. They do not consider the effectiveness of the text in teaching either reasoning skills or the philosophy of science. Nor do they report the reactions of their students to the text. This is not to say that philosophical adequacy is unimportant, only that it must be balanced against one's pedagogical aims. In responding to specific points, then, I am not merely defending what I wrote in USR. Nor is this the place for strictly philosophical debate. Mainly I want to point out the various considerations that led to the views developed. This should help others in understanding the book and in using it effectively. (It may also be of value to reviewers of other texts and to those considering writing a text of their own.)

Part I of USR contains a minimum of logical tools necessary to carry through the widely accepted philosophical view that one evaluates scientific claims by seeing whether they are *justified* by an appropriate *argument*. I felt it necessary to distinguish between deduction and induction, but not between logical (i.e., syntactic) truths and analytic truths. Whether the argument about the earth being larger than the moon requires an additional premise depends on whether the entailment is semantic or syntactic. Blurring this distinction, both syntax and semantics are part of form—the moon is subject matter. I doubt any students have been misled into thinking that affirming the antecedent is invalid because it might be regarded as an instance of the invalid form, P, Q, therefore, R.

To carry through the idea of justification by argument, it does seem necessary to introduce conditional statements. I did not, however, "neglect" to define the conditional. I deliberately *avoided* doing so. In particular, I avoided introducing the *material* conditional. Although many students find conditional thinking difficult, not many, in my experience, have been helped by being introduced to the material conditional. Nor is it necessary to give truth conditions for the case of a false antecedent since there are few scientific or practical contexts in which one could legitimately *justify* a conditional statement solely on grounds of having a false antecedent. Nevertheless, this is probably one place where I should have been more sensitive to the needs of instructors who have spent many years teaching propositional logic and therefore expect a more thorough treatment of the conditional relationship.

The amount of strictly deductive logic is limited because I am concerned to promote a type of course that is devoted primarily to *scientific* reasoning. Not that I would actively discourage others from using USR in a more traditional half deductive/half inductive course. I would only suggest that one be careful not to back into teaching such a course by uncritically adding material on deductive reasoning and then finding the term half over. One must plan carefully what to add, and then think about what parts of USR to cover. For a comprehensive and unified block that could be covered in seven or eight weeks, I would suggest the first two chapters of Part II (Reasoning About Theories) and all of Part III (Causes, Correlations and Statistical Reasoning). Alternatively, one might do Part III plus the last two chapters on decision theory (Part IV). If time is more limited, or if one wants to proceed at a more leisurely pace, one should do either Part II or Part III (omitting the decision theory), and not try to combine elements of both.

My use of a "definitional" (or "structural") as opposed to a "statement" view of scientific theories was not based on an interest in set-theoretical axiomatizations or in "the problem of theoretical terms." It was based on my belief that this approach provides a better account of what scientists actually do, and therefore a better framework for understanding what they are doing. And here I am particularly concerned with the biomedical and social sciences, where one finds much talk of "models" with a fairly limited range of application. But I now think I was wrong to *identify* theories themselves with definitions. This way of speaking conflicts too strongly with the way most people, including scientists, normally talk. In everyday speech, theories are things that may be empirically true or false, correct or mistaken, etc.

The view I was trying to capture can be better expressed by introducing the notion of a *theoretical model*. These are idealized systems (or types of systems) described by scientists. Such descriptions are best thought of as definitions. I would retain the idea of a theoretical *hypothesis* as an empirical statement to the effect that some real system is similar to the model in relevant respects and to an appropriate degree of accuracy. This emphasizes the scientifically important notions of idealization and approximation—notions that have been lacking in recent philosophy of science partly because they are difficult to capture in first order logic. *Theories*, then, are theoretical hypotheses of more or less general scope. Newton's theory is the claim that all the systems making up the heavens and the earth are Newtonian particle systems. Mendel's theory of genetics is the claim that many traits of organisms on Earth are transmitted according to some recognizable Mendelian model. In agreement with Gardner and Loewer's criticism, I would now make being deterministic or stochastic a property of *models*, not real systems.

In addition, following Kuhn, Lakatos, Laudan and others, I would now introduce the useful notion of a *theoretical tradition*. Thus specific theoretical hypotheses, but not theories and not theoretical traditions, can be fairly decisively refuted or confirmed by one or two good experiments.

Judging by my mail, the single most controversial thing in USR is Condition (2) for a good test of a theoretical hypothesis. As presented in USR, my two conditions for a good test of a theoretical hypothesis are:

1) If (H and AA and IC), then P.
2) If (Not-H and AA and IC), then very probably Not-P.

Here, H is the hypothesis being tested, AA are auxiliary assumptions, IC the initial conditions, and P a specific prediction. Now most experts agree that traditional accounts of the hypothetico-deductive method, represented by Condition (1), are incomplete. *Something* further is needed. There is, however, no consensus among the experts on just *what* further is needed. I am pleased that Gardner and Loewer recognize Condition (2) as an attempt to capture a widely shared intuition as to the *kind* of thing we all seek. I only wish they had spent a little more time seeing how it works in the classroom.

My reasons for formulating Conditions (1) and (2) as I did were primarily *pedagogical*. One will not find these formulations in my professional publications on probability and scientific inference. (See, for example, "Testing versus Information Models of Statistical Inference," in *Logic, Laws* and *Life*, R. G. Colodny, ed., Pittsburgh: University of Pittsburgh Press, 1977; or, better, "Testing Theoretical Hypotheses," in the forthcoming Volume X of Minnesota Studies, *Testing Scientific Theories*, John Earman, ed.) As noted above, I began with the philosophical position that evaluation of scientific information is to be analyzed in terms of justification by arguments. I then tried to limit myself to one basic form of argument, conditional arguments. This dictated that the conditions for a good test should be formulated explicitly as conditional, "If..., then..."

statements. I arrived at the specific form of Condition (2) by analogy with the corresponding condition for tests of *statistical* hypotheses which appears only much later in the book (Chapter 11). Long experience in teaching these materials convinced me that it is easier for students to begin with theoretical hypotheses because one cannot deal meaningfully with statistical hypotheses without introducing some more formal concepts of probability theory. It proved best to postpone introducing such concepts as long as possible.

There is, however, a disanalogy between statistical and theoretical hypotheses which leads to the kind of difficulty Gardner and Loewer point out. The negation of a statistical hypothesis may be quite well defined so that it is clear what the negation implies. The negation of a theoretical hypothesis is not so well defined. It is not clear, for example, what follows from the falsity of a Newtonian hypothesis. Thus the consequent of Condition (2) cannot be deduced from its antecedent. One has to bring other considerations to bear in order to decide, in any particular case, whether Condition (2) is satisfied or not. My choice was to try to build these considerations into the statement of Condition (2), or to keep the condition simple and introduce the supplementary considerations informally through the discussion of specific examples. I choose the latter course because I have found that students learn more from examples than from abstract formulas, particularly *complex* formulas.

In working through the examples discussed in Chapters 6 through 8, one learns the kind of considerations relevant to judging whether a particular prediction is likely to be successful if the hypothesis in question is false. These considerations are quite varied. One must ask whether there are any other good reasons—perhaps involving other justified hypotheses—for expecting the prediction to be true. Sometimes it is relevant to consider whether the hypothesis was explicitly formulated in order to account for the known or reasonably expected truth of the prediction. One's judgments in these matters are constrained mainly by current knowledge, not by the range of logically possible hypotheses that might yield the same prediction. This is in line with the suggestion reported in Gardner and Loewer's review.

Part of the reason for Gardner and Loewer's difficulties with Condition (2) is that they reinterpret it in a probabilistic framework alien to the framework of USR. My objections to their framework are not just philosophical. I don't see how one could possibly *teach* it effectively to the kind of students I wish to reach. Moreover, actual reports of scientific studies are rarely, if ever, formulated in such a framework. These are good reasons for not reinterpreting the conditions in this way.

Turning to Part III, if one is to understand reports on research in the biomedical and social sciences, one must be able to distinguish causation from correlation. One must also be able to distinguish kinds of *evidence* that support a conclusion of causation from those that only justify asserting a correlation. The scientific literature, unfortunately, is very circumspect on these issues. And philosophers have been primarily concerned with global analyses of the nature of causation, particularly with the adequacy of the Humean reduction of causation to constant conjunction. Standard philosophical treatments of the evidence for causal hypotheses rarely get beyond Mill's Methods. What I needed to accomplish the aims of USR was a simple account of the difference between correlation and causation that would tie in with the kinds of evidence that scientists in fact use in investigating these types of hypotheses. Such an account would make it possible for a nonspecialist to tell from standard sorts of reports whether or not a causal conclusion is justified.

The tendency of philosophers to think in terms of global analyses of causation has led many people to think that my account based on hypothetical populations was intended to be such a global analysis. This tendency was, unfortunately, reinforced by my insistence on distinguishing between the "meaning" of causal hypotheses and evidence for such hypotheses. But I never intended this account to be taken as a global analysis of the meaning of causation in general. It was intended as a restricted analysis that would provide a way of evaluating a usefully wide range of actual cases in the biomedical and social sciences. I knew full well that there are types of cases in which the analysis does not apply.

Indeed, in a discussion of the hypothesis that the use of marijuana causes heroin addiction (240), I explicitly stated that the real causal relationship may not be "correctly represented by the simple model developed in Section 9.5." And for just the reasons Gardner and Loewer bring up in their example of the hypothesis that homosexuality causes ostracism among humans. If everyone smoked marijuana, that would eliminate the intermediate link between using marijuana and the drug subculture in which the use of many illegal drugs is prevalent. Thus, if everyone smoked marijuana, the number of heroin addicts could be less than it is now. The other possibility mentioned by Gardner and Loewer, namely, that everyone in the actual population may already have the suspected causal factor, is of less interest because much less likely ever to be met in scientific practice. In any case, the fact that my simple model fails to be universally applicable does not justify the claim that I have made causation a contradictory notion. The real issue is whether the model is helpful in preparing students to analyze a reasonably wide and important range of cases.

USR does contain a global view of causality, but this view is not emphasized because it plays no role in the evaluation of causal claims. I included reference to this account in the section entitled "Causation in Individuals" mainly because I thought students would find it intuitively helpful. The view is that there are physical necessities operating in individual systems. I hold a similar view of probability, a propensity interpretation, of which USR contains not even a hint. I mention these more global views only because the fact I introduced causal necessities (though not by name) prior to developing the hypothetical population model shows that I did not think of this latter model as itself providing an overall analysis of causation.

I have never really been satisfied with Chapter 11, but not for the reasons Gardner and Loewer present. I think more time should be spent on interval estimation, which is a more useful notion than my treatment implies, and intuitively easier to understand than tests of statistical hypotheses. I also should have made more of the difference between the "logic" of tests and the methodology governing their use in specific scientific contexts. Finally, I should have illustrated the ideas with more realistic examples.

The scientific literature on hypothesis testing, like the philosophical literature on the hypothetico-deductive method, is unambiguous only on the *rejection* of the null hypothesis in the face of a statistically significant observed frequency. There is much less agreement on what follows from a non-significant frequency, though few methodologists would recommend the outright acceptance of the null hypothesis itself. My introduction of the "approximate null hypothesis," ΔH_o, thus goes somewhat beyond statistical practice. It has the pedagogical advantage, however, that it makes the response to statistical significance and non-significance as symmetrical as possible. And it exploits the analogy with the account of testing theoretical hypotheses developed in Part II. Gardner and Loewer's objection that I introduce the approximate null hypothesis primarily by means of graphs rather than giving an exact arithmetico-algebraic description only illustrates their insensitivity to the pedagogical aims of the book. No such description is possible without going way beyond the level of technical competence presupposed in the rest of the text. A similar comment applies to their later complaint that my chapters on decision theory contain no axioms for preference or utility.

The fact that the two conditional statements defining a good test are analytically true does have the consequence that any statistical test is a good test, but only in the "logical" sense that it is guaranteed (assuming a random sample of size n) to yield a *justified* conclusion. It is not guaranteed to yield a scientifically, or practically, *important* conclusion. That is the lesson of the sections entitled "Can a Good Test Be Not Good Enough?" and "Can a Good Test Be Too Good?" A justified conclusion may be of little scientific or practical importance, particularly when a small sample fails to yield a statistically significant difference. This is another place where the analogy between testing statistical hypotheses and testing theoretical hypotheses breaks down. Finally, I do regret that at some point the inequalities in Condition (2) got turned into equalities. That was a mistake, but not, I think, all that serious.

I am gratified that Gardner and Loewer find many strong points in Chapter 12 (Testing Causal Hypotheses). From the student's viewpoint, this chapter provides the reward for the effort expended in working through the sometimes difficult ideas in Chapters 9 through 11. In teaching Part III, I do not spend a great deal of time trying to get everything in Chapters 9 through 11 down pat. Rather, I proceed at a reasonable pace through these chapters so that everyone has a good familiarity with the ideas. Then I spend proportionately more time on Chapter 12, in which the previous ideas are applied to interesting cases. This gives students a chance to see the usefulness of these ideas and then to go back and pick up details they may have missed the first time through. The only drawback I have discovered in this strategy is that some students forget about statistical hypotheses or correlations, and try to analyze every study as a test of a causal hypothesis.

In introducing sampling distributions and statistically significant differences, it is much easier to treat the symmetrical case in which there are statistically significant frequencies on both sides of the frequency expected if the null hypothesis is true. But any observed statistically significant frequency will necessarily be on one side or the other. This invites one to draw an asymmetrical conclusion corresponding to the asymmetry in the data. There is a large methodological literature on the proper use of "one-tailed" rather than "two-tailed" tests. I avoid these difficulties by allowing the student to do what comes naturally and use the asymmetry in the data to draw the asymmetrical conclusion. To legitimate this practice one would have to introduce slight changes in the set of statistically significant frequencies or in the significance level. These changes, however, are not generally large enough to be worth treating in a detailed way, so I do not. Thus, contrary to Gardner and Loewer's charge, it is not really the case that I give "contradictory" characterizations of the null hypothesis. Rather, I allow a very natural additional inference that is not strictly justified in the framework presented, but which is close enough to correct to permit our avoiding the additional complexities a full justification would require.

There are general lessons in this exchange for everyone who writes, reviews, or uses textbooks. I hope they will be considered by all. □

Ronald N. Giere, History and Philosophy of Science, Indiana University, Bloomington, Indiana 47405, USA