Evaluation is the one indispensable activity of all science.

The Evaluation Taboo

Michael Scriven

Positivism, from its Comtean beginnings, developed as a backlash movement – against the worst excesses of German metaphysics, which at its best was vitually incomprehensible and at its worst was best left that way. In order to break this hold on serious metathought, members of the Vienna Circle, a discussion group that met at the University **of** Vienna from 1925 to **1936,** sharpened up Occam's Razor. In fact, they put an extra edge on it, applying it to concepts and not just entities, and laid about themselves with considerable vigor. It is hardly surprising that they over did it. It is more surprising that more than fifty years later we find leaders in Anglo-American-Australian-Scandinavian philosophy of science, such as Kuhn and Feyerabend, still cutting themselves on that razor, or running scared of it, and the leaders of the Continental school, such as Gadarner, Habermas, and Ricoeur, still trying to put the bits of the old metaphysics back together.

So much criticism of positivism has occurred to date, and there are so few defenders around, that it seems a little redundant to go into it any more. But careful study shows that much of the criticism is mere backlash itself, sometimes superficial or incorrect but nearly always unaccompanied by **a** constructive alternative, In many ways the most constructive alternative has been neo-positivism. Of course, such an approach has been criticized for being merely reformist, for not digging

E. R. House (Ed.). ***Philosophy of Evaluation.*** **New Directions** for Program **Evaluation, no. 19. San Francisco:Jossey-Bass. September 1983.**

deep enough to reconstruct a philosophical approach from a fresh foundation. But, in general, the aim is to build a structure that is at least razor-proof, though it will no doubt have its own weaknesses. There is a reasonable chance that such weaknesses will be less serious, since it is clear that philosophy of science, no less than science, accumulates theoretical strengths and knowledge over time—certainly across periods of half a century.

What is missing in the understanding of contemporary philosophy of science is virtually all work on normative science. In fact, there is scarcely ever any recognition of the necessity for normative science as a key category, and indeed its possibility is usually denied. Yet if one looks at the vast amount of recent material and professional activity in such areas as policy studies and program evaluation—the obvious examples—and begins to think about the nature of structural linguistics, medicine, educational or clinical psychology, welfare economics, quality control, and maintenance technology, it seems clear enough that a great deal of scientific work is devoted to determining how things *should* be and not how they are.

It is very difficult to explain this blindness to the evaluative aspects of present-day science without reference to the value-free concept inherited from the positivists—and I regard this gap in acknowledging values most significant of all the bad debts inherited from them. The blindness to evaluation is usually total: Not only is the possibility of any such macroentity as a normative science excluded, but also the possibility of scientifically established normative judgments within science, which would serve as microcomponents, is also not acknowledged. For example, in the standard listing of the conceptual activities involved in the nature of science—to be found in science as well as philosophy of science texts—one finds the routine mention of observation, measurement, concept formation, classification, generalization, inductive hypothesis generation (or speculation, if you prefer the Popperian view), prediction, hypothesis testing, and explanation. But there is no mention of evaluation. Yet without scientific evaluation—which can be defined as the systematic, objective determination of worth, merit, or other kinds of value—one could not use any of these processes in the service of science. The scientist's task is not to find just any explanation but the best explanation; not to use some experimental design but the best one that is feasible—and that one only if it is good enough to establish a worthwhile conclusion. The systematic and objective evaluation of the comparative and absolute merit of various entities, derived from the logic of science, is part of the basic skill repertoire of the scientist.

It is ironic that in the heyday of value-free science (that is, twenty years ago), it apparently never occurred to any professor in a science department that there was an inherent contradiction between the value-free doctrine and the claimed objectivity of the grades that were being given to students. Yet the process of evaluating student work in a science course reflects the process of evaluating scientific work in a scientific way—the very process that scientists are engaged in when doing science, either in reviewing the work of others or in thinking about their own work. There can be no sharp line drawn between the search for truth and the search for merit, since merit is one kind of truth and a key step on the way to finding almost all kinds of truth. If finding the correct answer to a problem is something that the pure search for truth is legitimately concerned with, then surely the correct answer to the problem of explaining something consists in finding the *best* explanation of that thing.

If *correct answer* or *true answer* (or *incorrect* or *false answer*) are not value-neutral terms, there is no conceptual base for a value-free search for truth. But if these terms are value-neutral, surely their logic can hardly be essentially different from that of *right answer* (or *wrong answer*)—both terms from the vocabulary of evaluation. *Best answer, bad answer, better answer, good reason,* and *bad inference* are other examples of evaluation terms. Scientists use this language in the search for truth and in the statement of their findings and could scarcely do without it, but it is by definition evaluative. Of course, it can be argued that all of these terms appear to be readily translatable, for all practical purposes, into the value-free language of description—of what is and is not the case. For example, good explanations of individual events are those that provide a set of true statements, including at least one law, from which one can validly infer the event that is said to be explained (according to the neo-positivist perspective). Such a positivist analysis contains no evaluative terms. So perhaps one should think of these uses of evaluation language as being dispensable, and the prohibition should be only essential uses of evaluation.

It is certainly true that there are legitimate central uses of evaluation in science. But there are plenty of uses of evaluation—such as description, explanation, and prediction—that are not legitimate, central, or even remotely scientific. That hardly shows these processes to be irrelevant to science. "Translatability" as used here is simply a test of respectability, and evaluative language passes the test just as well as explanatory, predictive, or descriptive language. More precisely, the proposed translation of *good explanation* is in fact a long way from the mark, since what it translates (and that not accurately) is *potentially good*

explanation. A good explanation must meet the further test of being better than any known alternative potential explanation. One cannot possibly argue that something continues to be a good explanation of some phenomenon if there is a better one around that is totally incompatible with it. However, evaluative terms inherent in that definition are self-evident.

The cases in which it appears that legitimate uses of evaluative terms are translatable into nonevaluative language are nearly always cases in which one can give fair translations in a particular context but not in general. Of course, one of the reasons for evaluation vocabulary is that it is transcontextual. Therefore, translation in context is not a way to show that it is dispensable. For related but not the same reasons, terms like *big* and *small* are extremely useful, even in quantitative science, despite the fact that we can replace them with something more precise in particular contexts. The very loose sense of *translation* that is advocated thus in the attempt to salvage the core uses of evaluative language is really equivalent to *substantiation:* essentially a reincarnation of the old positivist ploy about theoretical concepts. Some of the earlier positivists and neo-positivists argued that theoretical concepts are dispensable in favour of observational concepts, but what this often turned out to mean was that any particular use of the concept could be converted (translated) into a set of claims about the observable results from certain experiments. Not only was this false, because of the open texture of the concepts and the theories in which they were embedded, but it was also a bad reason for supposing that one could do without theoretical concepts. These were essential for transcontextual conversation about the nature of the universe or parts of it. All that had been asserted was that one ought to have reasons for claims about the presence of electrons, and that ultimately these reasons ought to be reducible to claims about the observable results of experiments.

In a parallel case, while it is true that one can give decisive factual reasons for evaluative conclusions in a particular context, for example, "the special theory of relativity is superior to the theory of elliptical waves," it does not follow that one can give translations, without a context, of evaluative terms into value-free language. Thus the fact-value distinction is really the same kind of logical point that leads us to say that theoretical concepts are not reducible to observational ones without loss of meaning, even though one often believes, and rightly so, that very good observational evidence can be given to support theoretical conclusions. However, this inability to reduce theoretical

language hardly provides a reason for abandoning theories. In fact, it is a very good reason for having them. Similarly, the inability to reduce evaluative language provides no reason for abandoning it, as long as it is clear and useful. Anyone who reads consumer product evaluations would agree to that.

Therefore, I maintain that evaluation is the only essential scientific activity. There are good examples of scientific work where one or the other of the usual list of acitivities is not present, but none without scientific evaluation. Without evaluation, done by scientists as a key part of science, one could not do or teach others to do science. What would it be like to do science or teach someone to do it if one were not allowed to distinguish between sound and unsound observations, helpful and unhelpful classifications, valid and invalid generalizations, valuable and trivial hypotheses, good and bad explanations, theories and instruments, inferences and experimental designs and ideas, important and trivial scientific papers, breakthroughs and run-of-the-mill scientific work, and, in the end, first and second rate scientists? And even if one learns to make those distinctions, how can one not evaluate those entities? In fact, the very distinction between science and pseudosciences such as astrology does not lie in subject matter (unlike the distinction between science and history, for example) but in the quality of the methods and inferences; it is essentially an evaluative distinction. Thus, science is itself, in this usage, an evaluative word, which means the body of *well-supported* knowledge about the world obtained as a result of scientific investigation, just as *the explanation* in one standard scientific usage is evaluative, which means the best or correct explanation.

Yet most scientists still will argue that science is value-free. Science is about as value-free as buying hogs or playing chess—and for the very good reason that any rational and reality-related, purposeful activity necessarily involves evaluation; it does not necessarily involve observation, measurement, prediction, or explanation. Then why are most scientists and many philosophers of science still attracted to a doctrine that is so strikingly incompatible with the evidence? Perhaps they are attracted to a doctrine that gives them an easy, although illicit, way of keeping out of the highly controversial debates about the social application of science—and also illicitly reinforces their perfectly legitimate claim to objectivity (in much of their work). There are legitimate ways to support both of these natural preferences on the part of scientists who do not happen to be in the normative areas of science, but the value-free doctrine is not one of them.

There is a deeper and more powerful force behind the avoidance of evaluation in science, which I call *value phobia.* This is the desire to avoid explicitly stated evaluation because of the threat, sometimes economic, that evaluation poses to others and oneself and hence the backlash that it generates. Professional evaluators—whether from the personnel, program, or even the product area—never cease to be amazed by the extent and severity of this threat and the extraordinary lengths to which people, including professionals and teachers of professionals, will go to avoid it. Of course, there is often a high price to be paid for avoiding evaluation, either by the society—as when teachers refuse to be tested for competency after reasonable doubt has been established that it is present—or by the individual—as when someone refuses to go to the doctor because the news may be bad. On the other hand, if evaluators believe in their own doctrine, they must be willing to check the costs and benefits of evaluation frequently and thoroughly, since it is quite clear that some types of evaluation cost a great deal more than they are worth and others are not worth doing at all if done poorly. In general, there are many problems in the philosophy of science that impact on the practice of science in serious ways. These problems are not easily escaped by scientists; with the exception of those more properly thought of as high-status mechanics rather than scientists. Few that dismiss it are able to avoid professional, pedagogical, or ethical blunders.

The radical left in the sixties also attacked what they called the doctrine of value-free science, but their reasoning was different from what is being argued here—and irrelevant to any doctrines that any serious scientist or philosopher of science has ever held. It would have been no surprise to Carnap, one of the leading logical positivists, to hear that scientists make value judgments when they select science as an area in which to work, or when they decide to apply it in the cause of the freedom (or dictatorship) of the proletariat. Those were exactly the kinds of value judgments that he would have given as typical of the human species—judgments characterized by the key fact that they cannot be scientifically supported as true or false claims. In fact, an explicit part of positivist doctrine is that science is simply a tool that can be used to good or bad ends, by anyone who chooses to use it. The radicals did not disprove the doctrine of value-free science; they simply asserted an already well-known consequence of that view.

However, in this discussion I am attacking the core of value-free doctrine by arguing that science is absolutely and essentially evaluative. It is an astonishing demonstration of the intellectual accep-

tance of positivism that the very term *value judgment* has come to mean a mere expression of taste or an unsupportable preference; but it has no such pejorative connotation in its literal meaning. Value judgments, like factual judgments and theoretical analyses, are of two kinds—the well-supported and the poorly supported. No scientist can avoid making them, although it is certainly possible to avoid making good ones.

The similarity between evaluation and theoretical speculation is more than accidental, since the value of something can be construed as a property of that thing in a kind of methodological space. Its value is a theoretical property of it within the theory of its function, nature, and use, like the way that the cause of particle events or space curvature is a theoretical concept within a given theory about the nature of matter or space. There is even a use of the term *theory* that is analogous to the cautious use of value judgment—the use in which it is distinguished from *fact*. But *theory* can be used in theoretical physics, where no imputation of doubt is involved. Such usage should be seen as corresponding to the use of the term *value* in program evaluation.

In general, the avoidance of the evaluative—the proscription of prescription—has done nothing but harm to science and society. It has led us into a fifty-year orgy of intellectually and socially pointless research, particularly in the social sciences, which was supposedly justified as "pure" research but could not possibly have survived serious evaluation. It has led universities into a lack of concern with the evaluation of themselves and their teaching within their institutions, which has cost all who have attended them dearly. In all of these matters, the radical left's instincts about the lack of social conscience in many scientists were right, and even their sense that scientists often take refuge behind phony doctrines—rationalizations—was also correct. But their attack was so irrelevant to the ideology that needed attack that it has had no lasting impact on the rotten core of thought at all. It tweaked a good many consciences but, as a result of the same kind of sloppy scholarship and easy rationalizing that led to the very doctrine that they were attacking, they missed the crucial target. However, attacking the value-free doctrine accurately and effectively is important because there are so many spurious reasons for adopting it and so much value-phobic pressure to accept them. For the philosophers of science such a value-free approach is an inexcusable example of shoddy work, with the counterexamples to be found in every scientific paper. But if the doctrine is not rendered completely absurd by complete exposure, it will simply continue to rise from the ashes.

Michael Scriven is a well-known evaluation theorist. He is currently professor of education at the University of Western Australia. His most recent works are Evaluation Thesaurus *(Edgepress, 1977) and* The **Logic of** Evaluation *(Edgepress, 1980).*