

Recent debates suggest that quantitative and qualitative approaches to evaluation are inevitably separated by a gap in worldviews. Critical multiplism is one way of thinking about evaluation that can unify these two approaches.

Critical Multiplism: A Research Strategy and Its Attendant Tactics

William R. Shadish

The quantitative-qualitative debate continues to split the house of evaluation. To their deepest adversaries, quantitative evaluators are all logical positivists ignorant of modern philosophy using outmoded methods. To their most steadfast opponents, qualitative evaluators are soft-headed radical constructivists who deny the concepts of reality and truth and so unwittingly deny the truth of their own approach. Despite the best efforts of those who have tried to unite the warring factions (Cook and Reichardt, 1979), such stereotyped interactions persist, and skirmishes between the interested parties regularly flare up into full-scale wars.

Central to the continued schism, I think, is the perception that quantitative and qualitative evaluators lack a common philosophical ground on which to meet. This chapter describes such a common ground, a unifying research strategy that all parties may be able to live with even if they do not endorse all its particulars. The strategy is not beyond criticism. In fact, criticism is one of its central tenets, so it welcomes criticism, and criticism is consistent with it. Nor will this strategy tell evaluators the right way to do things. In fact, it is premised on the notion that there rarely is one right way, so the best we can do is to try many ways while criticizing each. It is merely a way of moving us one step ahead of many current fruitless debates

This chapter is a substantial revision of a paper that appeared under the same title in L. Sechrest, H. Freeman, and A. Mulley (eds.), *Health Services Research Methodology: A Focus on AIDS*. DHHS Publication No. (PHS) 89-3439. Rockville, Md.: National Center for Health Services Research and Health Care Technology Assessment, Public Health Service, U.S. Department of Health and Human Services, 1989.

about paradigms. These debates are fruitless in no small part because of their continued obsession with a dead philosophy: logical positivism.

Logical Positivism Is Dead

It is fashionable to attack logical positivism these days, and it is easy to guess why: Logical positivism is dead, so one is sure to win. As Paul Meehl (1986, p. 315) pointed out, “logical positivism, in anything like the sense of Vienna in the late twenties, turned out not to be logically defensible, or even rigorously formulatable, by its adherents. It is epistemologically unsound from a variety of viewpoints (including ordinary language analysis); it is not an accurate picture of the structure of advanced sciences, such as physics; and it is grossly inadequate as a reconstruction of empirical history of science. So it is dead. All old surviving logical positivists agree, including my friend and teacher Feigl, who invented the phrase *logical positivism* and [who] was the first to introduce the approach in the United States in 1931. The last remaining defender of anything like logical positivism was Gustav Bergmann, who ceased to do so by the late 1940s.” Such disavowals ought to alert us that something is seriously wrong with any argument that treats logical positivism as if it were still a credible philosophical opponent. Put this way, many evaluators would agree.

But another argument is often forwarded to justify the continued focus on the flaws of logical positivism, namely that many evaluators or evaluation clients are implicit logical positivists. As one respected colleague put it in a recent conversation, what else would you call someone who believes that evaluation methods can yield definitive answers about the true state of affairs in a program? Admittedly, such a person could be called an implicit logical positivist if he or she assumed, for instance, that our methods can yield definitive answers. The problem with this reasoning, however, is that such assumptions are implicitly consistent with a host of approaches to science, not just with logical positivism. On closer examination, it makes much more sense to think of these assumptions as coming from some influence other than logical positivism, for two reasons. First, these implicit logical positivists almost never display knowledge of other aspects of logical positivism that are much more closely tied to the position. For instance, I have yet to see such a person endorse the use of predicate logic as a means of adjudicating theoretical disputes, although that is the logic in logical positivism (Bechtel, 1988). If they were logical positivists at all, surely they would know this. Second, such people do display knowledge that is more consistent with other understandings of science. For example, their emphasis on practical problem solving is sometimes more consistent with pragmatism than it is with logical positivism. Even more plausibly, perhaps their naive assumptions about science have sociological rather than philosophical origins. Specifically, since World War II,

modern science has marketed a highly idealized and naive image of itself to the public and the government. In that image, science is nearly infallible, always progressing, yielding ever increasing returns for each dollar allocated, and successfully self-monitoring and self-correcting. It seems far more plausible to attribute naive assumptions about science to our marketing of this naive image than to some presumed causal connection with logical positivism.

Defining the problem as logical positivism causes another difficulty. The definition of the problem contains the definition of the solution. If the problem is logical positivism, then the solution is educating evaluators and clients about logical positivism and about why it is such a bad idea. But if the problem is the marketing of a naive image of science, the solution is educating evaluators and clients about how science really works, warts and all. Viewed this way, the choice is clear. Educating clients about philosophy is, with few exceptions, a waste of our time and theirs.

For all these reasons, it is time to take on more worthy opponents than logical positivism. Fortunately, we do not lack for pretenders to the throne—social constructionism, naturalistic inquiry, hermeneutics, pragmatism, postmodernism, and a host of others. Each has staked out its turf, claiming more or less boldly to be first in line for the throne. But none has won wide intellectual acceptance, and none is heir apparent. The reason is that, while each of them may look pretty good when compared with logical positivism, all have significant problems when we compare them with one another—comparison that most tend to avoid.

The key contention of the present chapter is that there is good reason for this state of affairs: The main lesson of the last fifty years of epistemological and methodological debate is that no theory and no method provide a firm foundation for our inquiries. Hence, our guiding strategy should be to define an approach to inquiry that takes exactly this lesson as its main premise. The present chapter outlines such a strategy—critical multiplism (Cook, 1985). This strategy provides broad and general advice about how we should approach inquiry and about how we should select tactics to implement each of the many tasks that we face during inquiry. Some of these tasks include choice of methods, so methods are a special case of scientific tactics, but, because these tasks also include such matters as choice of question and decisions about how to facilitate the use of results in policy-making, this chapter will also show how critical multiplism approaches choice of tactics for these tasks as well.

Failure of Tactic

To understand why any particular tactic must be partly inadequate, consider the following question: What would the perfect method be like? The perfect method would have the following characteristics: It would be

so inexpensive that we could always afford it, even if our budget was small. It would be so robust that it could provide an exact answer to the question, even under circumstances that made it hard to implement. It would be so reliable that it yielded few if any systematic or random errors. It would be so feasible that any scientist could learn to use it quickly and accurately, and no scientist would be deterred from using it solely because it was too difficult. It would be so unobtrusive that it would not interfere with other tasks in the study by virtue of logical or logistical incompatibility. It would be so powerful that, when it was applied, everyone would concede the outcome. And finally, it would be so well conceived and justified that it would not become outdated as the result of subsequent developments that were more accurate or efficient.

Perfect methods or anything reasonably approximating them would be intrinsically attractive for many reasons. They would give us time in which to think creatively about solving scientific problems rather than obsessing about details of methodology. They would make science more efficient by providing quick and ready answers to practical problems in research design. And they would circumvent the need to learn about the wide array of methodological techniques that might otherwise be required across the many different situations that we face as researchers. In short, perfect methods would allow us to train scientists once and for all in how to do their work, so that they could get on with solving the important substantive problems that they faced without being troubled by questions of method.

But, of course, the search for perfect methods seems doomed to fail. A reasonable interpretation of the history of science is that we have yet to discover a single scientific method that has not proved to be flawed in some respect when compared with the criteria just outlined. This is true not only of social science but of the natural sciences as well. For example, perhaps the nearest thing to a perfect method ever invented was O. H. Lowry's method for protein measurement in biology (Lowry, Rosenbrough, Farr, and Randall, 1951). According to *Science Citation Index*, it was cited more than fifty thousand times between 1961 and 1975, five times more often than the second most-cited work (Garfield, 1979). Since the current best interpretation of citation counts (Garfield, 1979) is that they are generally acceptable measures of use, this high count clearly suggests that an enormous number of scientists have found Lowry's method to be the method of choice. But it has important flaws. The ideal protein measurement method would be optimally cheap, rapid, and accurate. Unfortunately, it is difficult to meet all three criteria at the same time. Tests that are rapid and accurate are expensive, tests that are cheap and accurate are slow, and tests that are cheap and rapid are inaccurate. Lowry's method comes closest of any method to being cheap, rapid, and accurate for protein measurement, which no doubt accounts for its ubiquity. But it, too, ultimately falls short. For example, it is not as accurate as other tests when

protein concentrations are small, and more recently developed methods have to be substituted in this circumstance.

If Lowry's method is not perfect, then can any method hope to be perfect? Probably not. Hence, while science must always develop better methods, such development may never provide a complete solution to the problem that single approaches are imperfect. The fundamental problem is that all scientific tactics are imperfect, because that is an inevitable feature of the finity or boundedness of human knowledge and action.

Failure of Strategy

However, the problem of methodological and tactical bias would not be anywhere nearly as severe as it is if scientists did not tend to adopt a systematically biased subset of these methods. But they do, because, as Kuhn (1970) has pointed out, scientists are taught from paradigms that include advice about the kinds of questions that are proper and the kinds of methods that are to be preferred. Each paradigm systematically includes some methods and excludes others. The result is that systematic biases of omission and commission are passed on to each new generation of scientists. To oversimplify, psychologists paradigmatically tend to prefer experimental methods and individual explanations of behavior, economists lean toward econometric models and profit maximization explanations, anthropologists sometimes gravitate toward case study and participant observation methods to construct a thick description of the case, and sociologists gravitate toward sociostructural explanations of phenomena. Compounding the problem, the social processes involved in teaching these paradigmatic strategies to novice scientists are more likely to exaggerate their superiority than they are to stress their strategic limitations. Consequently, we can expect that scientists will have significant difficulty conceptualizing their disciplinary strategies as biased rather than as best.

To some degree, of course, the narrowness of such paradigms reflects the fact that they are rightly tailored to the specific needs of the substantive area under study. For example, the double-blind, randomized clinical trial in medicine is a legitimate response to the needs of that field for answers about such questions as pharmaceutical efficacy. It is equally clear, however, that other paradigmatic methods are often unduly narrow or even misguided. For example, some psychologists rather mindlessly use some of the poorer quasi-experimental designs they learned in graduate school when the causal modeling techniques of sociologists or econometricians might be a more useful approach—or when causal questions should be abandoned entirely. Such problems, of course, become particularly acute when disciplinary paradigms are transferred en masse to interdisciplinary problems like program evaluation, where they are less clearly an appropriate response.

For all these reasons and others, scientists who rely on disciplinary training for scientific strategy are bound to produce research that is biased by paradigmatic errors of omission and commission. If so, then one of the more pressing needs in science is for the development of strategies that can uncover the biases of omission and commission that are inevitably present in all scientific methods and then ensure that they do not operate in the same direction in a study or in a research literature to yield a biased conclusion.

Critical Multiplism as a Research Strategy

Critical multiplism is a research strategy aimed at doing just that: It advises scientists to put together packages of imperfect methods and theories in a manner that minimizes constant biases (Cook, 1985; Houts, Cook, and Shadish, 1986; Shadish, 1986; Shadish, Cook, and Houts, 1986). Briefly, *multiplism* refers to the fact that any task in science can usually be conducted in any one of several ways, but in many cases no single way is known to be uniformly best. Under such circumstances, a multiplist advocates making heterogeneous those aspects of research about which uncertainty exists, so that the task is conducted in several different ways, each of which is subject to different biases. *Critical* refers to rational, empirical, and social efforts to identify the assumptions and biases present in the options chosen. Putting the two concepts together, we can say that the central tenet of critical multiplism is this: When it is not clear which of several defensible options for a scientific task is least biased, we should select more than one, so that our options reflect different biases, avoid constant biases, and leave no plausible bias overlooked. That multiple options yield similar results across operationalizations with different biases increases our confidence in the resulting knowledge. If different results occur when we do the task in different ways, then we have an empirical and conceptual problem to solve if we are to explain why this happened, but we are saved from a premature conclusion that a particular piece of knowledge is plausible.

Campbell and Fiske's (1959) multiple operationalism in measurement is the most widely known and appreciated form of multiplism. Its justification is prototypical of all multiplism, so it is instructive to review it here briefly. Any single measure contains bias of different sorts. For example, different investigators disagree about how a construct ought to be operationalized, and any single measure will undoubtedly reflect the preferences idiosyncratic to a small subset of all the investigators. Moreover, any single measure will contain some random error variance, some variance associated with the construct of interest, and some unique variance that is not related to the construct of interest but that is also not distributed randomly. Hence, no single measure is ever a perfect represen-

tation of a construct. Use of multiple measures helps investigators to note agreements and disagreements about how something ought to be measured, to note discrepancies in findings that vary as a function of presumptively significant differences among measures, and to compensate for the error and variance unique to any particular operationalization.

The move from multiple operationalism to critical multiplism is motivated partly by the observation that there is no logical reason for limiting it to measurement. Just as no single measure is perfect, no single approach to other tasks in science is without bias. Therefore, we may fruitfully explore the hypothesis that the same rationales that apply to multiple operationalism in measurement apply to the conduct of all tasks in science. That is, it might apply to question formation, where any single question is just one of many more or less important questions about a research problem, and no one question is always the right one to ask; theory or model selection, where any theory or model contains biases of omission and commission that make it at least partly incomplete or wrong; research design, where no research design is always optimal for answering a question, and under the best of circumstances each design feature clarifies different parts of a question; data analysis, where any particular way of analyzing data gives us an only partially complete picture of the results; interpretation of results, where many interpretations of the results of research will often have some validity; summarizing literatures, where we can never totally trust the results of a single study, but the aggregated results of multiple studies offer us an opportunity to explore how results vary over different ways of looking at a question; and research utilization, where no single method can ensure that the results of research will be useful or used either by other scientists or by the community that shapes policy. Each of these propositions represents a specific instance of the general strategy represented by critical multiplism: Always seek alternatives that can shed a different light on the problem.

Guidelines for Critical Multiplists

If I were to list guidelines for an approach to research from the critical multiplist perspective, I would have to formulate two kinds of guidelines: technical guidelines and social guidelines. Both kinds of guidelines would apply to all seven tasks outlined in the preceding paragraph.

Technical Guidelines for Critical Multiplism. The first kind of guideline is technical. There are seven technical guidelines:

1. Identify the tasks to be done.
2. Identify different options for doing each task.
3. Identify the strengths, biases, and assumptions associated with each option.
4. When it is not clear which of several defensible options for doing a task

- is least biased, select more than one to reflect different biases, avoid constant biases, and overlook only the least plausible biases.
5. Note convergences of results over options with different biases.
 6. Explain differences of results yielded by options with different biases.
 7. Publicly defend any decisions to leave a task homogeneous.

The technical tasks specified by these guidelines aim at creating heterogeneous biases in planned research and identifying the biases that were and were not present in completed research.

We can illustrate these guidelines first with an example from basic research in psychology, then with one from applied research on the topic of AIDS. In the former case, personality theory has long discussed a controversy referred to as the *person-situation debate*. Essentially, it involves the extent to which individual behavior is influenced by person variables, such as personality traits, or by characteristics of the situation, or by some interaction between the two. Not too long ago, I and two colleagues applied critical multiplist guidelines to an analysis of the research bearing on that debate (Houts, Cook, and Shadish, 1986). We began with the premise that many options that might be studied were implicit in the question—many personality traits, many behaviors that might serve as dependent variables, many kinds of situations, and many data analytic strategies that might provide indexes of behavioral consistency. Various combinations of these options constituted different ways of framing the person-situation issue. Through a review of the literature, we were able to show that researchers in this area had examined a remarkably small combination of the possible questions that could have been asked. More important, individual researchers often framed the issue in different ways. When their research seemed to generate different results, it was often because they had asked systematically different versions of the question. In the few cases where they asked the same version of the question, they tended to get the same answer. Finally, we were also able to uncover some biases of commission in these studies, errors yielding results that were plausibly biased in the direction that we believed the authors to have preferred, given their theoretical predilections. I will return to this example often as this article proceeds in order to elaborate how critical multiplism can be applied.

Now consider an example from AIDS research, a pressing applied social research issue. Suppose that we wished to answer the following question: Are high-risk groups changing their behavior in a way that might minimize their likelihood of being infected? The next step, if we follow critical multiplist guidelines, is to identify the options for answering this question, options that might plausibly result in different answers, and then try to ensure that no important option has been systematically overlooked in the literature as a whole. The answer to the question may depend, for

example, on the group that we elect to study: homosexuals, intravenous drug users, or hemophiliacs. If we are studying a treatment, the answer may depend on whether that treatment is administered in individual counseling or to a group of peers in which peer pressure can be brought to bear, or it may depend on whether treatment is administered by health professionals or by community organizations that historically have access to and credibility with the target population. The answer may also depend on the dependent variables that we elect to examine: knowledge of AIDS risks, actual safe-sex practices, sharing of needles, or promiscuous heterosexual contact. It may depend on whether we assess the occurrence of these behaviors by simple self-report, by observing behavior where feasible, by assessing gonorrhea infections and hepatitis as proxies for the likelihood of AIDS infection, or by trying random response techniques. It may depend on whether we collect these dependent variables in the long term or the short term, or cross-sectionally or longitudinally, and on whether we have a federal certificate of confidentiality with which we can at least try to reassure respondents. Finally, it may depend on whether we conduct the study in San Francisco, Indianapolis, or Nigeria. All possible combinations of these options suggest the universe of studies that could be conducted to answer this question. Clearly, no single one of these versions of the question is the correct one, so our research strategy should be to ensure that we have overlooked no plausible version of the question unless we can make a good case that it would not yield results different from the research that has already been done. Undoubtedly, a review of the AIDS literature would reveal that some of these versions of the question had been studied extensively, while others had not been studied at all, and that the answer to the question is positive in some versions but negative in others. Organizing such a review around these multiple options would point to future studies that seem important to do but that have not yet been done, and it would force us to try to explain why the answers to the question vary depending on how the question is operationalized.

This example is not meant to imply that critical multiplism limits itself to options for formulating good research questions. It is much broader than that. In the example just given, we could have focused on design options for answering a treatment outcome study based on this question. We could have focused on options for estimating whether the results of the study would generalize to the population of high-risk persons in the United States. We could even have focused on whether this question was worth asking at all given other competing questions that could be asked about, say, the prevalence of seropositivity or the costs of AIDS. In all these matters, there will be different ways of proceeding—different designs, analyses, and questions to ask—all of which we need to consider carefully before we draw conclusions about AIDS that might be premature until the complete set of options has been thoughtfully considered.

Being Multiplist with Limited Resources. Resource constraints usually prevent us from including all the heterogeneous options that our analysis suggests are warranted. Hence, we must often decide what to make multiple in a particular study. The following corollary guidelines can help us to make the choice:

1. Include only the options that can be defended as presumptively likely to yield different results.
2. Give preference to options that have been left homogeneous in past studies.
3. Decide how much uncertainty reduction is desired or needed given the situation, and then select options that might be able to provide that level.
4. Select options that are financially feasible in the context of the study. Use the following rough guides to help conceptualize the costs of making different options heterogeneous: The least expensive options include multiple measures on the same subjects, multiple data analyses by the same investigator with different models and assumptions, and the submission of research proposals and reports to multiple critics. The moderately expensive options involve multiple kinds of subjects, observers, and occasions; multiple analyses of the same data set by different investigators; and the hiring of multiple consultants for on-site visits. The most expensive option is implementation by multiple independent investigators at different sites.
5. Publicly defend any decisions to leave a task homogeneous.

Particularly for options that are presumptively important in their implications for changing the results of a study, there is little excuse for not implementing at least the least expensive options for being multiplist.

Social Guidelines for Critical Multiplism. Critical multiplism requires the investigator to be aware of all the tasks that need to be done, the options for doing each task, and the biases and assumptions associated with each option. But individual scientists are limited in their capacity to know all these matters, they have difficulty perceiving their own biases, and they are insulated from obtaining some of the knowledge that they need by such sociological structures as academic disciplines (Faust, 1984; Mahoney, 1976). The problem is compounded because, as D'Espagnat (1983, p. 133) put it, "the human mind is spontaneously overconfident on such matters. In particular, it too frequently raises ideas that look 'clear and distinct' to the level of absolute truths. It sometimes fails to note the cases in which the notions in question are, in fact, of questionable validity or are relevant merely within some limited context."

For such reasons, we must expect that individual investigators will fail to complete the technical tasks of critical multiplism in a way that clearly

identifies all the options and biases. Thus, we must have a second set of guidelines aimed at remedying the social and psychological limitations of individual investigators in accomplishing the technical component in a satisfactory manner:

1. Identify several sources (people, past research, competing theories) whose biases are likely to differ from those of the person who completed the technical tasks.
2. Enlist the aid of those sources in completing and criticizing the results of the technical component.

In AIDS research, for example, this social psychological component of critical multiplism could be implemented by seeking input from multiple groups that have an interest in AIDS so that we can be more confident that important perspectives and options have not been overlooked; commissioning multiple independent studies of the same question by different investigators around the nation, each of whom would be encouraged to communicate but to pursue independent investigations of the topic; submitting raw data from all these studies to multiple analysts, each of whom would use different assumptions and techniques to uncover biases of omission and commission; and sending the report of results to multiple interest groups who could point out hidden biases and assumptions in the study and its interpretation. In all these examples, a cardinal guideline is that criticism is fostered best when its source has a perspective that is quite different from that of the original investigator.

What Critical Multiplism Is Not. It might help us to clarify critical multiplism if we consider what it is not. The opposite of multiplism is monism—always doing some task in the same way. The opposite of being critical is to be mindless in the choices that one makes. Hence, the extreme opposite of critical multiplism is mindless monism—always doing things in the same way without thinking about the strengths and weaknesses of one's choices. This is probably more prevalent than we would like to think, partly because biases in disciplinary training tend to expose us to limited sets of options and partly because thinking critically about one's work is both intellectually and emotionally demanding. Nevertheless, being a mindless monist is probably not a good thing. It is a surefire way of getting into trouble, because good critics will eventually find the problems with your research unless you find them first.

However, we should also consider the other two possible options, mindless multiplism and critical monism, for they afford more interesting contrasts. Mindless multiplism is introducing some variation in options without really thinking much about it. My favorite example is most uses of stepwise multiple regression. Stepwise multiple regression throws multiple predictors into a regression; the computer selects the ones to enter

into the equation. Very little in science is more mindless than this, and in general such work is to be discouraged. However, we might be willing to allow at least a minor place for truly mindless multiplism in research in the interests of generating diverse and unanticipated perspectives on a problem, much as Campbell (1974) called in his evolutionary epistemology for blind variations in the social problem-solving options that we choose to study. It might maximize the chance of encountering some truly novel solution to a problem. Problematically, Campbell's justification really requires the variations to be random, and mindlessness seldom achieves that goal, because its tendency is to fall back on habit. Habit is not random.

However, there is a place in research for critical monism—doing many things in the same way even though one has thought critically about the possible weaknesses of doing so. The tendency toward critical monism results from the fact that our time, skills, and resources are limited, so we can only implement one or at most a few options in any given study. Still, examining the strengths and weaknesses of the options that you do choose critically will help you both to choose the best option for your study and to anticipate valid criticisms of the options that you have chosen so you can qualify your conclusions accordingly. Hence, it is more realistic to strive for some mixture of critical monism and critical multiplism in one's research. I discuss these practical matters again at the end of this chapter.

Applications of Critical Multiplism

Earlier, I listed seven tasks that scientists must address: question formation, theory or model selection, research design, data analysis, interpretation of results, summarizing literatures, and research utilization. The examples in this section show how critical multiplism can be applied to each of these tasks.

Critical Multiplism and Question Formation. Here, the critical multiplist searches for constant biases in the kinds of questions that have and have not been asked and for bias in the way in which past research has operationalized questions. Critical multiplism in question formation is particularly important because the questions that scientists ask implicitly contain both a conceptualization of the problem that they are addressing and the seeds of actions that might be taken to remedy the problem (Rein and Schon, 1977). Thus, without critical scrutiny of these questions, scientists risk a wrong or biased conceptualization of the problem and consequent ineffective or insufficiently limited action. Unfortunately, we probably know less about how scientists formulate their questions than about any other aspect of their behavior. This omission is probably all the more unfortunate since the value of research may depend more on the value of the question being investigated than on any other single factor, because the value of the enterprise as a whole can be compromised from the

start by investigating the wrong question. Playing on statistical terminology describing Type I and Type II errors in testing the null hypothesis, Dunn (1982) refers to this problem as *making Type III errors*—setting confidence intervals correctly when testing null hypotheses about the wrong problem.

However, in critical multiplist terms, the problem is not one of simply asking either the right or the wrong question. Rather, it is that we should be wary, either as individual researchers or across a field as a whole, of asking a possibly good set of questions that nonetheless systematically shapes the issues in a biased manner. Thus, we should carefully scrutinize and criticize our questions on two levels. The first level concerns the general kinds of questions that we are and are not asking. For instance, in the early part of this decade, federal agencies seemed to be reluctant to fund AIDS research that dealt with the question of safe sex. I know of one research team whose grant proposal on that topic had conveniently been lost by the agency for a prolonged period of time. The investigators were eventually given an indirect message that such research was not being encouraged because it implicitly condoned homosexual behavior, which was not consistent with the political climate then current at the federal level. Fortunately, this grant proposal was eventually found and later funded, speaking either to the power of a life-threatening crisis to influence the political process or, as one observer put it, to the more mundane but equally powerful influence of the condom lobby, given the profits that it might make if some safe-sex habits were adopted (Turner, 1988). If nothing else, this example underlines the extent to which question formation is a social, political, and economic process and so must be critically analyzed as such.

But we should be more worried about instances of biased question formation that are less obvious than the example just given. A colleague of mine suggested that we should be studying the long-term social psychological impacts of AIDS on how we relate to one another in American society. His concern stemmed from a conversation with his eight-year-old daughter, who came home from school discussing not only the explicit details of sexual activity but also her fears of such activities, since they might ultimately have fatal consequences. She had not learned all this from any AIDS education classes, since such classes had not yet been implemented in the public school system that she attended. Rather she gleaned most of it from her peers—eloquent testimony to the way in which social and sexual norms are being challenged and shaped among young children. We may be living with this consequence of the AIDS epidemic long after we have achieved some success in limiting its physical consequences.

Another example of a gap in the kinds of questions that we ask about AIDS concerns women and AIDS. Probably largely because women account for such a small percentage of AIDS cases, we have very little data on such

matters as female sexual behaviors regarding AIDS or the course of the disease in women. But looking down the road ten years, we may regret this omission if AIDS spreads increasingly to the heterosexual community, if we begin to be more concerned about the spread of AIDS among prostitutes, or if the issue of abortion rights confronts the issue of bearing babies with AIDS.

Once we have decided to ask a particular kind of question, there are still several different ways of asking the question that might plausibly yield different results. We have already examined the many different ways in which we can address the question of whether high-risk groups are changing their behavior. Another example comes from the colleague who works on the aforementioned safe-sex study. In tracing the epidemiology of homosexual behaviors so as to help identify safe-sex practices, the investigators in that project made the initially plausible assumption that an increasing number of sexual partners was the key factor in becoming infected. Therefore, they neglected to ask about the number of encounters with each partner, a variable that later research proved to be an important factor. Significantly, colleagues at the same institution had already determined that the number of encounters was important, but my colleague's team did not discover this owing to the lack of communication across teams at the institution—a point that underscores why the social psychological component of critical multiplism is so important.

Another example concerns the examination by Houts, Cook, and Shadish (1986) of the person-situation debate. That topic had been marked by decades of controversy and seemingly contradictory research about whether personality traits, situational constraints, or some interaction between person and situation best predicted behavior both within and across occasions and settings. Yet our analysis showed that many of the differences among authors resulted from the fact that they were asking quite different versions of the same question. The most important source of difference between thinkers concerns their sampling of traits, behaviors, settings, and occasions. Epstein (1979) used multiple behaviors reported over multiple occasions to examine multiple traits, but the behaviors often occurred in a single setting or in multiple settings that could be unconfounded with occasions. Hence, Mischel and Peake (1982) considered his work to have limited relevance for the issue of transsituational stability. Buss and Craik (1983) dealt with multiple settings, multiple traits, and multiple behaviors within each trait category. However, their work is less relevant to concerns about the situational stability of a particular discrete behavior (behavioral consistency) than it is to concerns about dispositions defined as measures of different behaviors presumed to belong to the same latent trait (dispositional consistency). Both Bem and Allen (1974) and Mischel and Peake (1982) examined two traits, multiple behaviors within each trait, multiple settings, and multiple occasions.

However, Bem and Allen (1974) do this in an idiographic mode, so that only Mischel and Peake (1982) have independently sampled traits, behaviors, settings, and times in the nomothetic mode in which the person-situation debate has been mostly pursued. Sampling has to occur within each of the four factors if analysts are to explore the stability of behaviors across settings, especially when the behaviors are presumed to represent a trait, and we want to develop theory that is not specific to a few idiosyncratically selected traits. It follows that the person-situation debate has rarely been about the same questions, so that much of the disagreement has been about defining the question, not necessarily about the answer to the question that has been posed.

Critical Multiplism and Theory or Model Selection. Here, the task is to ensure that inadvertent constant biases have not crept into our selection of the theories and models that we use to conceptualize a problem and guide our research on it. Any theory or model contains biases of omission and commission that make it at least partly incomplete or wrong. This statement is true even in the most well-developed physical sciences. For example, the theory of quantum mechanics applied to particle physics may be the single most successful theory in the history of science, because it integrates and explains an enormous array of diverse observations. Yet even here, physicists and philosophers disagree about how to explain certain experimental phenomena that seem to imply a counterintuitive nonseparability of distant particles (D'Espagnat, 1983; Rae, 1986). The interviews with physicists involved in this controversy presented in a recent, highly accessible book (Davies and Brown, 1986) make it clear that they propose radically different and inconsistent explanations of the problem, none of which seems to be completely successful. In the social and medical sciences, of course, we lack theories or models as well developed as quantum mechanics. But there are still plenty of examples that point vividly to the need to retain and consider multiple theories and models. For example, in learning theory in the 1950s, Spence and Tolman debated the role of cognition in learning, with Tolman's position receiving significant vindication only twenty years after the debate was thought to have been resolved in Spence's favor (Gholson and Barker, 1985). In the medical sciences, many molecular biologists and geneticists rejected Barbara McClintock's work on mobile genetic elements because it was inconsistent with constancy of the genome, the dogma that prevailed at the time (Keller, 1983; Lewin, 1983). Rejected theories and models have a way of coming back to haunt us in newer, better-supported forms that force us to revise our theories and models once again. Disconcertingly, these matters are seldom superficially clear, and it is rarely that one theory is widely thought to be best even for a short time. Far more often, theories and models contend actively, each backed by a group of scientists that fervently argues on behalf of its pet approach.

Since AIDS research is a relatively new endeavor, it is even more apparent that we are still groping to locate the best theories and models for it. It is probably fair to say that we view AIDS mostly as a medical and psychological problem, so we apply available medical and psychological theories and models to it. Just as it is true that the way in which we frame questions shapes the actions that we later recommend, so it is true that the theories we bring to bear shape the actions that we take. If we view AIDS as a medical problem, we recommend medical and public health solutions. If we view it as a psychological problem, we see the solution as lying in changed attitudes and behaviors. Other theoretical perspectives on AIDS are far less common. But perhaps we should consider them as well. For example, blacks and Hispanics account for about 40 percent of the men with AIDS, although together they represent only about 17 percent of the population. The overwhelming majority of women and children with AIDS are members of minority groups. If we explore this overrepresentation of minorities, we might well find that economic and social causes are as critical or more critical than biological and psychological ones. That is, it may be that AIDS is more prevalent in these populations than in others for the same economic and social factors that contribute to drug abuse, prostitution, and other phenomena observed in the underclass. If so, we would be led to consider social and economic solutions to the problem. Of course, critical analysis might then lead us to conclude that we cannot do much about the root social and economic causes of drug abuse or prostitution and that it is therefore more feasible to implement discrete behavioral solutions, such as distributing clean needles. But we cannot do this critical analysis until we have at least considered some alternative conceptualizations of the problem. And in the years to come, if medical and behavioral solutions seem insufficient to cope with AIDS, we may be forced to reconsider some social and economic interventions.

Metatheoretical differences also had an effect on the evolution of the person-situation debate. Some researchers—especially Bem (1983)—preferred an idiographic approach that left subjects free to define which behaviors represented which traits. Others—especially Epstein (1979)—operated within a more traditional nomothetic conception that defined traits uniformly over people. Still others—especially Mischel (1979)—operated within a more social-cognitive theory that replaced traits with information-processing concepts or—like Buss and Craik (1984) in particular—operated within a theory of personality that conceptualized traits only in terms of overt behaviors assumed to indicate common latent traits. These metatheoretical differences led each thinker to formulate and study different questions, as we saw in the preceding section.

Critical Multiplism and Research Design. In designing research, we often find that we could use any of several different methods to construct an answer. When many different options have been used to study a

problem and when they all converge on the same result, we are tempted to be confident in the result. But such confidence is misplaced if all these methods nonetheless still share a common bias. One example concerns evaluations of federal social programs aimed at improving pregnancy outcome—such programs as the Maternal and Child Health Program (MCH) or the Special Supplemental Food Program for Women, Infants, and Children (WIC). One review that examined about thirty such evaluations (Shadish and Reis, 1984) found that the studies used a wide array of methods, including surveys, one-group pretest-posttest designs, regression modeling, and time series. Across these different methods, the result was largely the same, namely that the programs improved pregnancy outcomes. Yet despite the apparent diversity of method, the studies all shared a common bias. That is, almost none of the studies reported whether the mother had received treatment from a program other than the program under review, although it seemed quite likely that mothers were receiving additional services, since there are more than seventy federal programs that can seemingly have an impact on pregnancy outcomes. In fact, the one study that did document mothers who had received other treatment (Sharpe and Wetherbee, 1980) found that the addition of WIC services to the other treatment had no beneficial effects.

For an example from AIDS research, we can examine models of the spread of the HIV infection. Turner, Fay, and Widdus (1988) reviewed three different models of this problem that yielded estimates ranging between 750,000 and 2.5 million infected persons. The models were multiplist in the sense that each model used different assumptions, different sources of data, and different statistical procedures. Turner, Fay, and Widdus (1988) added a commendably critical analysis of these different models, locating significant sources of uncertainty in each. Thus, they pointed out that one model depended on a fraction that changes substantially as an epidemic progresses, so that estimates from a patient population in which the virus has only recently been introduced will differ substantially from estimates for populations in which the virus has only recently appeared. Another method used estimates of the number of homosexual males in the United States compiled in the 1940s by Kinsey, Pomeroy, and Martin (1948) from data gathered among college-educated people in the midwest—data whose relevance to today's problem is doubtful.

But from a critical multiplist perspective, the analysis just reviewed still lacks one important feature. In all their otherwise excellent discussion, the authors did not attempt to estimate the direction of the biases introduced by these uncertainties. For example, experts in the field undoubtedly have some knowledge of whether the Kinsey data over- or underestimate the prevalence of homosexuality among American males today. Armed with such knowledge, we would then be in a better position to judge whether the 1986 estimates produced by the U.S. Public Health

Service were likely to over- or underestimate the number of persons infected with the HIV virus. More generally, if we could determine the likely direction of bias in each of the three models, we would be closer to understanding the accuracy of current estimates.

Dawes (1989) suggests a somewhat different example in his discussion of rating scale methods of assessing the risk of various behaviors that can lead to AIDS. He points out that all the questions on the AIDS knowledge and attitude surveys conducted by the National Center for Health Statistics (NCHS) use traditional normative rating scale response formats. In fact, every such survey ever done on this topic uses normative rather than ipsative formats (Turner, 1988). Dawes (1989) points out that the ambiguity in such rating scales makes it very unclear exactly how risky the public understands these behaviors to be, especially relative to one another. He suggests an alternative ipsative methodology, ranking or paired comparisons techniques, that would eliminate much of this ambiguity and thus yield more accurate assessments of our understanding of the relative riskiness of various behaviors.

To turn again to the person-situation debate (Houts, Cook, and Shadish, 1986), method selection among the many researchers evolved naturally to include at least some multiplism. Over studies and less rarely within studies, researchers have examined multiple personality traits, multiple situations, multiple occasions, and multiple criterion behaviors. But this evolution was haphazard at best, and there was little careful thought about how one could best sample multiple options with each of these four categories. For example, selection of traits was guided more by convenience and prior substantive interest than by a theoretical analysis of traits in general or of the types of traits that might be more or less consistent across situations. The same thing can be said for sampling of situations, sampling of behaviors, and the time frames under which cross-situational consistency could be expected to be observed. Consequently, the findings that accumulated were of quite uncertain generalizability. Perhaps most seriously, when the major research question is one about behavioral stability across situations, it is crucial to examine the same behavior in different settings, yet none of the studies examined did any systematic unconfounding of situation with behavior. As a result, the key question of cross-situational consistency remained quite open.

Critical Multiplism and Data Analysis. To begin with, a healthy dose of exploratory data analysis has no equal for creating heterogeneity of perspective (Tukey, 1977). Moreover, for almost every data set, there are several analytic methods that could be used to answer a question. Sometimes there is evidence that one analytic technique is probably superior to others, as when modern linear structural modeling techniques replace analysis of covariance (ANCOVA) to adjust for group nonequivalence given the problems that ANCOVA incurs as the result of unreliable

measurement (Cook and Campbell, 1979). Often there is doubt that one analytic technique is best, in which case more than one technique ought to be used so that we can see whether varying our analytic methods causes our results to vary. In such cases, we can test competing causal models, as several authors did with the Head Start quasi-experimental outcome data (Cicirelli and Associates, 1969; Magidson, 1977; Rindskopf, 1981). Rossi, Berk, and Lenihan (1980) provide another example. In analyzing data from a randomized experiment, they used both traditional analysis of variance that tested the direct effects of treatment and three-stage least squares to investigate some mediating processes. In each of these cases, more was learned from the joint use of multiple analyses than from any single analysis in isolation.

For an example in AIDS research, we can look at a recent protocol from the Agency for Health Care Policy and Research (AHCPR) projecting the number and cost of future AIDS cases as a function of the many variables over which the estimate could change. The AHCPR notes that such variables include the extent of seropositivity in different areas of the country, the probability of HIV infection associated with different modes of transmission, specific cofactors that can increase the likelihood of transmission for a given individual, the duration of the incubation period, and a patient's expected life span after acquiring ARC or AIDS. The AHCPR protocol calls for flexible mathematical models incorporating these key parameters that permit us to perform sensitivity analyses in order to determine the parameters that contribute most to the uncertainty of future projections. Thompson and Meyer (1989), using different assumptions about parameters that might affect estimates of the costs of AIDS, advocate similar sensitivity analyses, and Lagakos (1989) suggests that survival analysis can be used to give a range of plausible estimates that realistically portray the uncertainties encountered in understanding the course of AIDS.

The analysis of the person-situation debate by Houts, Cook, and Shadish (1986) provides another example of critical multiplism in data analysis. There, we reanalyzed a data set that Mischel and Peake (1982) had published. We were guided by the hypothesis that several biases in the data may have led those authors systematically to underestimate the overall cross-situational consistency in behavior: The reliability of the items measuring traits was generally quite low, which would attenuate correlations. The reliability of behaviors was also low. The authors had not used a Fisher z -transformation in cumulating correlations (while this transformation is controversial, its use would have increased correlations even further). Moreover, we had reason to think that item distributions might have been skewed or restricted in range, which would further lower correlations. Finally, the procedures that the authors had used to average over items may have underestimated cross-situational consistency because

the items may not have all been good measures of the same core construct. While our reanalysis could not remedy all these problems, the corrections that we could apply indeed supported the suspicion that problems in the original analysis may have systematically underestimated cross-situational consistency. This case shows that multiple analytic options can lead to biases that all run in a constant direction.

Critical Multiplism and Interpretation of Results. The goal in applying critical multiplism to the interpretation of research results is first to generate an array of possible interpretations, then to assess those interpretations as either more or less plausible on the basis of available evidence. In AIDS research, a particularly ingenious example was provided by Valdiserri and others (1988), who were concerned with interpreting the fact that the HIV virus can be recovered from saliva (Groopman and others, 1984). Some might interpret this finding to mean that intimate kissing is not a safe sexual practice, since the virus could be transmitted through exchange of saliva. But by bringing in multiple lines of evidence that provided new information from different perspectives, Valdiserri and others (1988) developed another interpretation. That is, since recovery of the HIV virus from saliva has been shown to be far less frequent than its recovery from blood (Ho and others, 1985), intimate kissing is less likely than other sexual activities to be a likely route of transmission. Findings from studies of a related infection support this hypothesis. That is, it has been shown that oral-oral contact is not related to seropositivity for the hepatitis B virus although that virus is found in saliva (Schreeder and others, 1982). While there was probably still some possibility of transmission through intimate kissing, Valdiserri and others (1988) interpreted this marginal possibility as allowing intimate kissing to be classified as a “possibly safe-sex” practice.

We see another example in the reanalysis by Houts, Cook, and Shadish (1986) of the Mischel and Peake (1982) data on the person-situation debate. Mischel and Peake (1982) asked whether the trait of conscientiousness predicted cross-situational consistency. Somewhat haphazard procedures and criteria had been used to generate the items pertaining to conscientiousness. Consequently, different observers could have different opinions about the items that were appropriate measures of the trait. Hence, we used only the items that subjects considered to be most typical of conscientiousness, and the results again suggested more cross-situational consistency than the original authors had reported. Differences in interpretation about what counts as conscientiousness thus lead to differences in conclusions about the stability of behavior.

More generally, we suggested that the results of research on the person-situation debate could be interpreted from the vantage point of similar work in social psychology on the relations between attitudes and behaviors. This area, like the other, has generated large amounts of

research and controversy, and researchers in both areas have spent a great deal of time trying to agree on estimations of the magnitude of the relationships between traits and behavior on the one hand and between attitudes and behavior on the other. Yet the literature on attitudes and behaviors moved much more rapidly to address two other kinds of questions: questions about the contingencies that govern variation in magnitude and questions about the mechanisms through which behaviors express general attitudes in specific situations. This clearly pointed to a gap in the person-situation literature and allowed us to describe an extensive program of research that might help fill it.

Critical Multiplism and Single Studies. For all the reasons discussed on the preceding pages, we can never totally trust the results of a single study. So if a single study reports that HIV-infected teenagers developed AIDS more slowly than either infants or adults, we are best advised to note the finding but view it with suspicion. But when several studies conducted by independent investigators in different cities all report the same finding, we should note the convergence of results. Even so, we should still look for constant biases that might plausibly account for the finding. Indeed, all three studies had a constant bias—all were conducted with hemophiliac teenagers. While we have no reason to believe that AIDS progresses differently in the presence of hemophilia, it may be that the longer latency observed in these studies is nonetheless an artifact of the fact that hemophiliacs as a group may have been infected with HIV somewhat more recently (for example, 1982–1984) than homosexuals (1978–1980), given that some time is bound to have elapsed between the initial appearance of the virus and contamination of the blood supply used by hemophiliacs. Unfortunately, this is one case that gets more rather than less complicated as multiple lines of evidence are brought to bear. That is, other lines of evidence support the finding that children may develop some diseases more slowly or less severely—chickenpox or measles, for example.

In AIDS research, Turner (1988) analyzed public perceptions and behavior in response to AIDS by examining, searching for convergences and divergences, and in some cases reanalyzing data from four different national surveys. Thus, he found that public fear of AIDS varied as the result of how the question was asked. Between 17 and 28 percent of the respondents answered in the affirmative when they were asked whether they were afraid or concerned about contracting AIDS, but only 10 percent—about half as many—said they felt that they were at risk of contracting AIDS. Similar variations were found in public attitudes about quarantine for AIDS victims, with only 17 percent saying that they should be quarantined as lepers are but with 51 percent agreeing that AIDS should be added to the list of diseases that must be quarantined. We can almost always learn more from a critical analysis of multiple studies than we can from examination of the results of a single study.

The analysis of the person-situation debate by Houts, Cook, and Shadish (1986) reflects the same principle. It was the contrast among the many studies that we reviewed that caused us to appreciate the variety of dimensions on which the studies differed—traits, behaviors, situations, and occasions, to name only a few. With this larger perspective, we could then see the gaps in each individual study, which in turn led us to guide our reanalysis of the Mischel and Peake (1982) data partly by trying to study those gaps. But by cumulating studies, we could also examine the findings that were common despite the fact that the individual studies were so different in so many ways. For example, studies generally agreed that personality tests are of little use in making predictions about single behaviors in specific situations and that there was strong evidence for the temporal stability of behavior within situations.

Critical Multiplism and Research Utilization. No single method will suffice to ensure that the results of research will be useful or used, either by other scientists or by the community that shapes policy. Those involved in applied social and medical research learned this lesson the hard way over the last twenty years (Shadish, Cook, and Leviton, 1991). From simple reliance on such traditional techniques of dissemination as publication and presentation at conventions, scientists who want their work to be used have added to their repertoire such techniques as preparing executive summaries outlining major findings and recommendations in simple terms, consulting with users before starting the research to learn what kinds of information they want and might be able to use, keeping in close contact with users during the course of the study, providing interim results prior to the final report, and making use of media presentations in such forms as news conferences and press releases.

By virtue of their training and socialization, however, most scientists probably avoid most of the techniques just identified except those tied to publication and interaction with peers. AIDS research is probably somewhat of an exception to this rule, given the extent of public concern about the problems involved. Media attention has been forced on many AIDS researchers, and especially at the federal level, the nature of federal funding for AIDS research tends to focus researchers' attention on gathering data that federal agencies see as holding the most promise for immediate use. But we might still wonder what kinds of biases have crept into our conceptions of useful AIDS research. While I am generally not familiar enough with AIDS research to answer this question with certainty, I can say that I would ask two questions about this matter. The first concerns biases in the selection of stakeholder groups—the groups that, directly or indirectly, are affected by AIDS and AIDS research—whose input has helped to shape the kinds of information that AIDS researchers have gathered. We have probably done a good job of gathering data that might be useful to the general medical and public health community and to the federal agencies

most directly concerned with treatment for AIDS patients and paying for AIDS treatments. But it is not clear that we are trying as hard to provide useful information to the dying AIDS patient who wants to know where and how he can most comfortably live out his remaining time in dignity, to the local school superintendent who wants to know how to convince angry parents that it is in the community's best interests to educate their children about AIDS, or to the families of hemophiliac children who do not know how to cope with their child's illness and who see AIDS destroying their hopes, dreams, and finances.

The second question that I would ask concerns the distinction between short-term and long-term use. For very understandable reasons, AIDS research has adopted a crisis mentality, with nearly all the federal AIDS dollars going into research aimed at coping with the immediate impacts of the AIDS crisis. But there is another kind of use that we should also be thinking of, a kind of long-term use aimed at enlightening us about the full scope of the AIDS problem in all its ramifications and about the long-term impacts and changes that it may have on American society. What, for example, would John Naisbitt, author of *Megatrends* (Naisbitt, 1984), say about the way in which AIDS will transform our lives? This kind of research, broader and perhaps more exploratory in scope than research aimed at solving more immediate concerns, is nonetheless just as essential to our long-term ability to cope with AIDS.

Invalid Objections to Critical Multiplism

Some readers might raise three objections to critical multiplism: First, there really are methods in science that are uniformly acknowledged to be superior—for instance, the randomized experiment and the random sample survey in social and medical research. As a result, critical multiplism is misguided and even misleading in suggesting otherwise. Second, the tenets of critical multiplism cannot be true because, when we apply them reflexively to critical multiplism itself, they imply that it, too, is imperfect and cannot lay claim to being the right approach to inquiry. Third, critical multiplism implicitly approves the notion that anything goes and considers all methods, practices, and interpretations to be equally valid. None of these objections is accurate. But let us consider each of them in turn.

Random Assignment and Sampling as Ideal Methods. Most readers need no introduction to the ideas of random sampling from populations and of random assignment to experimental conditions. Both ideas have a long and illustrious career in science (Campbell and Stanley, 1963; Fisher, 1925; Rossi, Wright, and Anderson, 1983). Random sampling helps enormously with generalization from samples to the populations from which the samples have been drawn, and random assignment to conditions greatly facilitates causal inferences. There can be no question that these are

two of the most useful methodological tools in the social scientist's repertoire.

The first thing to note, however, is that both techniques are in fact multiplist. Recall that multiplism aims to make heterogeneous those aspects of research that might bias results. The aim is to leave no systematic bias dominant in the results. In principle, both random assignment and random sampling accomplish exactly this with regard to the unit of assignment or sampling, making it likely within a known probability that no systematic biases are present either among units in the sample or across groups. In research where the sampling unit is a person, as is often the case in social science and medical research, this amounts to treating people as a facet of the research design and making subjects heterogeneous so that they do not share any irrelevant characteristic that might be confounded with the desired inference. Random procedures are therefore the most plausible ways of accomplishing the goals of multiplism, especially because they can be taught routinely to scientists and implemented robustly in research.

This point underscores that the key guideline of critical multiplism is not that multiple options should always be implemented. Rather, the key guideline is hedged by the condition *when it is not clear which of several defensible options for a scientific task is least biased*. Only then should multiple options with divergent biases be selected. It is sometimes the case that an option is in fact quite plausibly the least biased of several options that could be implemented. Since that will often be the case for random procedures, they will often be the method of choice.

But for many reasons we are likely to overestimate the value of random procedures and to place confidence in them in situations where it is not warranted. Our overconfidence is probably rooted in the social and psychological processes involved in teaching the techniques to novice scientists during their initial disciplinary training. Because we want to encourage new scientists to think routinely about such methods and because time limits what a single course during graduate school can cover, we often simplify the presentation of random techniques and put off discussions of their weaknesses and vulnerable assumptions to a later date. However, this later date often never arrives, so that scientists then tend to use these methods in rote fashion without much further thought. After all, the pressures of a research career often make us reluctant to take time to reflect on possible flaws in the methods that we habitually use or to learn new techniques that are both conceptually and methodologically difficult. Under these circumstances, routine reliance on random procedures may not be such a bad thing if it serves to ensure that many scientists are using methods that, on the whole, are among the best we have.

Nonetheless, criticizing random procedures reminds us of their many limitations and encourages us to think of alternatives in the many situa-

tions in which random procedures are not the best choice. To begin with, random procedures work well only if certain assumptions are met. When those assumptions are not met, as they often are not, random procedures must be complemented or even replaced by other methods. For example, consider random assignment to conditions. Chen and Rossi (1987) claim that random assignment to conditions facilitates causal inference only to the extent that the sample is large, or the sample is homogeneous, or the experiment has been repeated many times; there are no interactions between treatment variables and extraneous variables; attrition rates are the same among experimentals and controls; and conditioning variables affect all important subdivisions of the experiment evenly.

Many of these same assumptions, or variables on them, also apply to random sampling. Thus, Cook and Campbell (1979) point out that random assignment to conditions does not control for the problems in causal inference that occur when treatment intended for one group is communicated to another group (treatment diffusion), when program administrators cannot tolerate the focused inequality that results from restricting treatment to some subjects and so offer compensatory treatment to subjects in control or less-desirable-treatment groups (compensatory equalization), when subjects assigned to less-desirable treatments try to show that they can do just as well as those in treatment groups (compensatory rivalry), or when subjects in no-treatment control groups become demoralized and do less well than they otherwise would have (resentful demoralization).

Of course, one might take issue with any one of these problems. For example, it seems likely that one of Chen and Rossi's (1987) conditions—that there be no interactions between extraneous variables and treatment—is less a problem for molar causal inference than it is for the construct validity of treatment. And their first three conditions are less a problem for the logic of randomization than they are for the interpretation of individual experiments in which one or more conditions are not met. Nevertheless, all these points are well taken, at least in some respects. In fact, these problems seem all the more formidable because they are frequent in field research, such as AIDS research, where we often encounter small samples, heterogeneous samples, unreplicated experiments, nonrandom or differential attrition, important interactions, subjects who communicate with one another, administrators or caregivers who will not stand by and see clients assigned to treatments they think will not work, and patients who react emotionally and behaviorally about the treatment or control condition to which they have been assigned.

These are not just hypothetical problems. On the contrary, they are often encountered in field research. For example, in AIDS research, Valdisseri and others (1987) found that homosexual and bisexual men who agreed to participate in an AIDS education session were more likely to have

had a college education than those who did not respond. Since better-educated people also tend to have more accurate knowledge about virtually all aspects of AIDS (National Center for Health Statistics, 1988), the study effectively sampled subjects who needed the intervention least and who probably represented only a minority of the population at risk. Similarly, AIDS researchers assume that infected patients who are assigned to treatment conditions will actively seek treatment from other sources, thus compromising the integrity of clinical trials. The implication is that, even if random procedures are often theoretically superior to other methods, they will encounter problems that have no easy solution and that thus make them biased in ways that we often cannot adjust for. Of course, we can always choose to ignore the presence of unknown but plausibly systematic biases on the grounds that this is the best we can do given current technology. But those grounds are defensible only to the extent that the researcher has just one possible method. However, the only thing that limits us to a single method in research is habit, disciplinary training, or simple narrow-mindedness.

The criticisms of random procedures just outlined are not the important ones, for they really argue only that we should complement our use of random procedures with some other procedures that might be able to compensate for their weaknesses. Random procedures still occupy center stage in such a scenario. But they begin to move to the side when we consider more serious objections that question the ability of random sampling or random assignment to bear some of the burdens that we encounter in field research. Two objections seem particularly important. First, even knowledgeable proponents of random sampling and random assignment recognize that they were not designed to address certain important research problems. Returning to the distinction between strategy and methods, we see that both random sampling and random assignment are matters of research method, not of research strategy. Since methods should always be subservient to strategy, it follows that the researcher must first and foremost be concerned with such strategic matters as the kind of question to be asked; the way in which the research project is to be implemented; how the data are to be analyzed; and how the results are to be interpreted and synthesized with the results of other studies that can shed some light on the same question. Neither random sampling nor random assignment have much to do with most of these strategic matters, but critical multiplism does.

When such procedures as randomization are placed in this context, it is easier to see how truly exceptional they are. They are exceptional both in the sense that they are unusually useful and accurate when we can apply them successfully and in the sense that they are the exception rather than the rule. That is, for most matters of research strategy, we do not have methods uniformly acknowledged to be superior even in principle. This is

especially true for such matters as the generation of research questions or the interpretation of results, where we have very little understanding of how scientists generate options and choose among them. It is almost as true for the synthesis of research over multiple studies, where, for example, the fledgling methods of quantitative literature reviews are still the center of great controversy. It is even somewhat true for data analysis, for there are often several interesting and valid ways of analyzing the same data that yield different perspectives on it, a point that Coombs (1964) highlighted in his seminal work on measurement theory, *A Theory of Data*. In all these strategic matters, no single method is least biased. As a result, we should consider and implement multiple options, each having different biases.

The second serious objection to random procedures is that they are often not feasible in field settings. For example, consider random assignment to conditions. To be sure, such procedures are often far more feasible than they are said to be (Boruch, 1975; Boruch, McSweeney, and Soderstrom, 1978). We can even grant that such procedures may be significantly underutilized in field research and still be correct in asserting that there will be many important situations in which they are not feasible for reasons of logic, time, resources, skills, or logistics. Under these circumstances, there most often is no single best fallback position for facilitating causal inference and no single tactic that yields a least-biased answer. As a result, the best research strategy is to implement multiple options that do not share a single bias. For example, when randomized experiments are not feasible—and excepting these rare instances in which some of the strong quasi-experimental alternatives like time series or regression discontinuity can be used—there is often no single best tactic that yields a least-biased estimate of cause-and-effect relationships. The two major options—the various kinds of causal modeling and the use of quasi-experimental design features—both fail in different ways to deal with the problem adequately. Quasi-experimental design features are aimed at identifying the presence of threats to the validity of a causal inference, but they often can provide only limited estimates of the magnitude of causal effects, and they usually cannot eliminate the effects of many biases that they find to be present. Causal modeling techniques tend to rely on analysis of correlational data to adjust away bias and estimate the magnitude of effects, but their assumptions about our ability to know such biases completely and measure them perfectly are questionable. When, despite the different strengths and weaknesses of these techniques, the joint use of causal modeling and quasi-experimentation yields a convergent answer, our confidence in the causal inference is increased. It is even better if we implement each of these options in a multiplist mode, testing multiple causal models that vary in their assumptions about selection biases or the predictors of outcome and including in our models multiple quasi-experimental design features (multiple pretests, nonequivalent dependent variables, cohort controls),

each selected to shed some light on a plausible bias (Shadish, Cook, and Houts, 1986).

Similar arguments apply to the feasibility and appropriateness of random sampling. Such sampling is most often successful when we sample from populations of persons or similar units, such as classrooms or hospital wards, for purposes of generalizing back to those populations. Yet there are many important research situations in which this is either not feasible or inappropriate to the generalization being sought. The most obvious situations are those in which the population at issue is not easily identified. The population of persons with AIDS or of persons susceptible to AIDS is an excellent example. In such situations, it may be more feasible to use alternatives to formal sampling from populations, such as sampling of presumed heterogeneous instances and of presumed modal instances (Cook and Campbell, 1979). Of course, we could always use a national random sample survey to estimate the magnitude of this population, but it would be certain from the outset that this survey would encounter important biasing feasibility problems (Turner, Fay, and Widdus, 1988). For example, if we used a household-based sampling frame, we would overlook some important groups, such as intravenous drug users, who are not attached to households but who are particularly susceptible to infection. No matter what the national sampling frame, some degree of nonresponse correlated with seropositivity is likely to occur, making estimates of this already low base rate characteristic all the more biased. A national random sampling study cannot solve all these problems. One excellent suggestion for dealing with this problem is to sample from respondents to a recent national survey, such as the Health Interview Survey, so that data from its interviews can be tapped in analyses of the biases that nonresponse may have caused (Turner, Fay, and Widdus, 1988). This is critical multiplism at its best, using multiple tactics that plausibly help to estimate and compensate for the biases in various sources.

Less obvious threats to the superiority of random sampling are the many situations in which we want to generalize but use a variety of tactics that have little or nothing to do with random sampling. In test construction, for example, modern formulations of classical test theory—domain sampling theory (Lord and Novick, 1968; Nunnally, 1978) or its extension in generalizability theory (Cronbach, Gleser, Nanda, and Rajaratnam, 1972)—allude to a sampling of items that in fact never takes place. Rather, generalization from the items to the construct of interest is justified on the basis of such tactics as the matching of item content with an explicated theory of the construct and on the basis of statistical covariations among items that make it plausible to think that they measure the same thing.

In still other situations, the notion of sampling to achieve generalization seems to disappear entirely. Consider, for example, the notion of generalization of causal connections (Cook, 1990). Cronbach (1982)

approaches this task by seeking detailed explanations of the mediating links and mechanisms among inputs, causal processes, and outcomes, each of which is specified at a level of detail sufficient to clarify the essential mechanisms in each that are necessary to achieve the effect. According to this notion, generalization is facilitated by the fact that this detailed knowledge of mechanisms allows us to identify the essential elements, which we can transfer to new sites with increased confidence that the effect will be reproduced. Cronbach, in short, seeks generalization through theory, not sampling. Another approach, best represented by meta-analysis, seeks causal generalization through identification of robust causal connections that occur over a heterogeneous array of persons, settings, times, and cause-and-effect constructs. Such generalization is supported by two rationales. One rationale is induction—an effect observed over heterogeneous instances is more likely to appear in future instances not represented by any study in the current meta-analysis. The other rationale is falsification—the scientist deliberately seeks to find exceptions to the causal connection by varying instances within the existing data set. In meta-analysis, neither of these rationales relies explicitly on sampling or on Cronbach's notion of theory. If we examine all these instances of causal generalization in detail, we discover a multiplicity of tactics, each of which has advantages and disadvantages varying with the situation and none of which can uniformly be preferred to others over the many kinds of situations that scientists face (Cook, 1990).

In summary, then, it is probably wrong to object to critical multiplism on grounds that such procedures as random assignment and random sampling are uniformly superior. Not only is it wrong, it is also unfortunate, because such an objection suggests that random procedures somehow conflict with critical multiplism. Nothing could be further from the truth. Not only are random procedures in keeping with the best multiplist assumptions about making bias heterogeneous, but such procedures are no more than a small, special, tactical case of the general strategic approach represented by critical multiplism. Confusing tactics with strategy only clouds the issues involved.

The Reflexivity Problem. Does critical multiplism apply to itself? Yes. To be true to the themes of this chapter, therefore, there is one sense in which I must discourage widespread adoption of critical multiplism. After all, because no approach to inquiry is perfect, multiplism dictates that we encourage multiple approaches to be tried and then criticized. Indeed, such proliferation will inevitably continue as advocates of other positions develop them to their fullest and show what they can do that critical multiplism cannot do. The community of interested scholars will compare and contrast the strengths and weaknesses of each approach to find its biases of omission and commission. Eventually, our sense of the benefits of each approach will improve.

Of course, this very process is critically multiplist. That is one of the beauties of critical multiplism. Its advice applies perfectly well to itself. Those who follow the larger literature on science studies may recognize how hard it is to achieve such reflexivity without self-contradiction. For example, a key tenet of some social constructivist positions, especially the more radically deconstructionist ones, is that all knowledge is merely social construction and that no particular piece of knowledge is more valid than another. When applied reflexively, of course, this means that the approach called social constructivism is itself just a construction, hence no more valid than any other approach. At best, such a conclusion somewhat moderates the force of the position as a whole. At worst, it refutes it entirely. Discussions of this point among social constructivists never resolve the problem. Essentially, they argue that the benefits of deconstruction outweigh any problems caused by self-contradiction (Ashmore, 1989; Woolgar, 1988). Critical multiplism gives us the benefits without the contradiction. It encourages deconstruction, even of itself, without endorsing the nihilist extremes. It accepts the lack of firm foundations for knowledge without rejecting the possibility that some constructions of knowledge are better than others. Of course, the last point raises problems about how one distinguishes better from worse constructions, a point to which I now turn.

Does Critical Multiplism Mean Anything Goes? Another kind of objection is that critical multiplism flirts with a total rejection of standards for what is to count as a better or worse approach to the doing of science. Put as a simple question, can critical multiplism be distinguished from philosopher Paul Feyerabend's (1975, p. 28) speculation about whether "anything goes"? This criticism is particularly astute because the notion that anything goes is similar to critical multiplism in some less than obvious epistemological ways. In particular, while it is always difficult to trace the influences of other people's work on the development of an idea, it is probably accurate to suggest that one intellectual root of multiplism is Campbell's (1974) notion of blind variations in evolutionary epistemology. He draws an analogy between the way in which species evolve biologically and the way in which knowledge develops in science and other areas. Central to both kinds of evolution is the availability of radically different alternatives that help to cope with a problem in a truly new and perhaps more effective way. The argument is that an alternative is less likely to cope with a problem differently precisely in the degree to which it resembles what already exists. Hence, truly novel variations are needed. In biological evolution, these variations are accomplished by means of random genetic mutation. By analogy, evolutionary epistemology requires a similar mechanism. The analogy begins to break down at this point, because we cannot say what mechanism serves this random permutations function in science. But some version of multiplism is a likely candidate.

One feature distinguishes this multiplism from anything goes: criticism. Multiplism encourages us to explore diverse options. But such exploration is not an end in itself. Rather, each option must be evaluated for its ability to solve problems in the relevant research. In essence, we must ask four questions of each option: First, will a bias of omission or commission be incurred if we do not implement the option? Second, will the option reduce that bias better than other options now available, or will it at least reduce bias below current levels? Third, can we estimate the direction and magnitude of these biases for the option? Fourth, if multiple biases are associated with the option, can we combine our assessments of their direction and magnitude in a way that enables us to make an overall judgment about the value of the option?

In operationalizing this logic, we must keep in mind that, although many biases are possible, not all are plausible. In critical multiplism, plausible biases always take priority over possible biases that are not obviously plausible. This corollary is necessary in order to avoid an infinite regress that allows anything goes to return because we cannot exclude any possible bias, no matter how implausible it may be. Thus, any operationalization of these four questions is plausible precisely to the degree that it is not inconsistent with such criteria as reason, experience, or observation. How plausible a bias is determines the priority that we must assign to its study. Ultimately, of course, these judgments of plausibility are as fallible as other kinds of human knowledge, and so they will be subject to revision and reinterpretation over time (Lakatos, 1978). We may thus have to conceptualize plausibility in Bayesian terms, constantly adjusting our estimated plausibilities in light of new evidence. Sometimes we may not know exactly in which direction to make these adjustments, especially if there are several plausible biases, each of which may operate in different directions to produce either a positive or a negative effect (Mark and Shotland, 1985). Even when the scientist has absolutely no indication which bias is more plausible, he or she is better off for having taken the critical multiplist approach because identifying plausible biases points to questions for future research.

Valid Objections to Critical Multiplism

Certain problems need further attention before critical multiplism can be as practical as required. The pages that follow describe some of these difficulties and the directions that possible remedies might take.

What Is Bias? The term *bias* figures prominently in the present discussion. Yet I have deliberately left it ambiguous, primarily because what counts as a bias varies with the different tasks that scientists must perform. Statistics is the context in which the notion of bias is clearest: Bias implies some systematic (that is, nonrandom) distortion so that the value

of a statistic is no longer “true.” Statisticians speak of *biased* estimators, *biased* samples, and *biased* tests. They can also speak of, say, linear *unbiased* estimators and asymptotically *unbiased* estimators. In this sense, too, we may say that random samples yield asymptotically *unbiased* estimates of their population parameters or that randomized experiments with no attrition yield *unbiased* estimates of treatment effects. In these last expressions, *bias* has a highly stylized and formal meaning.

Unfortunately, most scientific tasks are not as clear-cut as statistics, so the meaning of *bias* changes accordingly. However, in most uses it retains the notion that some systematic phenomenon interferes with the inferences that we want to make, but it loses the notion that some “true” value is being distorted. For example, if we say that a certain researcher is biased against a certain theory, we mean that his or her visible behaviors favor one theory over another, even if we do not know which theory is true—or even whether a theory is true. Similarly, if we say that some literature is biased by its failure to examine a certain perspective, we do not need to know whether that perspective is true in order to note that the literature is biased.

In general, then, bias is an error of omission or commission that leaves us either with knowledge that is in some plausible respect wrong or misleading or without certain kinds of knowledge that are plausibly needed. For example, the general failure in AIDS research to ask questions about women and AIDS may leave us without information that we may need to cope with future extensions of the AIDS epidemic into the heterosexual community. Our failure to ask about the socioeconomic causes of AIDS among prostitutes may deprive us of a significant option for treating the disease in that population. Our failure to use ipsative measurement strategies like rank ordering and paired comparison techniques in research design may cost us accurate knowledge about the general public’s perceptions of the relative risks in the various ways in which AIDS can be transmitted. These are all plausible biases of omission or commission.

Bias is also particularistic and fallible. It is particularistic in the sense that the arguments both for and against it necessarily have to appeal to particular, specialized knowledge that is germane to the research topic at hand. For example, in justifying why one thinks that randomized experiments are more conducive to causal inference than other methods, we must appeal to concepts and methods involving the nature of causation, the nature of statistical probability, and our inability to specify selection biases with the degree of precision that we would like. But our arguments will also be fallible, because they are based on assumptions about such things as causation, probability theory, and the magnitude of selection biases that may be incorrect in some way; because they require method to be implemented with a theoretical fidelity that practice often belies; and because they are always subject to revision as new observations and theories render them obsolete. Therefore, exploration of the nature of bias preferentially

occurs at the particularistic, not the general, level. I have provided some examples of how we can make such applications in the context of past work on quasi-experimentation (Shadish, Cook, and Houts, 1986), long-term care for the chronically mentally ill (Shadish, 1986), and the person-situation debate (Houts, Cook, and Shadish, 1986).

Thus, what we need are ways to explore, define, and estimate such biases across a variety of domains. At least four directions seem promising in this regard. The first is Monte Carlo modeling. For example, Zwick and Velicer (1986) describe five different ways of determining the number of factors to extract from a correlation matrix during a factor analysis. They found that the eigenvalue-greater-than-one rule systematically tends to extract too many factors from a matrix. While Monte Carlo methods have most often been applied to statistical problems, it seems that we could easily use them to examine a wider array of methodological and design problems:

A second way of estimating biases is by adapting meta-analysis to the task (Glass, McGaw, and Smith, 1981; Hedges and Olkin, 1985). Meta-analysis has mostly been used to review studies for the substantive conclusions that they yield about a topic, as in the numerous meta-analyses of whether psychotherapy works (Smith, Glass, and Miller, 1980). But a growing number of scientists are interested in the use of meta-analysis to examine the effects of tactics and methodologies on the results of research. Most prominent have been the many examinations of whether randomized experiments and quasi-experiments yield different effects (Smith, Glass, and Miller, 1980). Another example is empirical studies of the biases in published studies relative to so-called file-drawer (that is, unpublished) studies (Shadish, Doherty, and Montgomery, 1989). Meta-analysis seems to have enormous potential for improving our empirical understanding of the tactical biases in research.

The third way of estimating bias is to manipulate particular tactical options experimentally to see whether they produce different results. Much of this kind of work has been done in survey research. For example, Bradburn and others (1979) were interested in investigating biases that might result from the various ways in which threatening questions could be asked. They randomly assigned respondents to one of four conditions: personal interview, telephone survey, self-administered questionnaire, or the random response method. They found both that "no one method is clearly superior to all others" and that "the random response method does appear to produce less underreporting for very threatening questions" (Bradburn and others, 1979, p. 167).

The fourth way is through better observation, description, and measurement of the methodological features of research about which we currently know very little. Examples include Haveman's (1985) description of the costs and benefits of social experiments; LaLonde's (1986)

description of empirical differences between randomized experiments and selection bias modeling in estimations of the effects of social interventions; and Latour and Woolgar's (1979) description of the ways in which laboratory scientists form questions, shape data, and interpret results. Especially in this last area—we know so little about how scientists proceed—discovery-oriented methods may be most appropriate for our first attempts at exploring the biases that may be at work.

While the volume of research on bias seems to be increasing rapidly, there is still far too little of it, and it tends to be scattered across literatures and mostly to be limited to traditional statistical and methodological matters. If we are to have a truly practical critical multiplism, we must have more such studies, their implications for the conduct of research have to be better integrated, and—perhaps most important—they need to be extended to such relatively overlooked but critical tactical matters as formation of questions, interpretation of results, integration of findings over multiple studies, and dissemination and use of scientific knowledge.

Biases That Everyone Shares. Another problem with critical multiplism is that when all available theories or methods share the same bias, even multiplistic criticism may be incapable of revealing it. This criticism is a version of the notion shared by contemporary philosophers of science that we can never really know that we have uncovered the “truth” (Lakatos, 1978; Laudan, 1977). While this is a serious problem, it does not mean that critical multiplism is less worthy than other strategies for the conduct of science, because shared biases have proved to be a major stumbling block in all science. In fact, I know of no technique for ensuring that all such bias has been avoided, so I will grant that all knowledge is presumptive rather than “true.” I would only hypothesize that critical multiplism is more likely than other strategies to uncover these shared biases when they exist.

Social Psychological Resistance. One important problem with critical multiplism is that efforts to implement it are likely to meet social and psychological resistance from scientists. Recall that the need for the social psychological component of critical multiplism arose from scientists' inability to identify the biases in their own work. Some of these limitations are psychological. For example, even with the best intentions, human beings are finite in their capacity to see all the alternative questions, designs, analyses, and interpretations that could characterize a study (Faust, 1984). Failing to see those alternatives, they tend to reach premature closure that an issue is settled. Another psychological limitation involves the biases that seem to be a part of human nature. That is, all scientists are human beings who seem to be subject to many of the same psychological and social biases that we can observe among other human beings (Faust, 1984; Mahoney, 1976). Like all humans, scientists are prone to ignore whether their predictions exceed base rates, to draw unwarranted generalizations from a small number of observations, to misinterpret

regression artifacts as real causal agency, and to pay attention to things that confirm their favorite theories while passing over things that refute them (Faust, 1984). Both scientists (Mahoney, 1976; Mitroff and Fitzgerald, 1977) and nonscientists (Bradley, 1981) tend to be passionately committed to their favorite hypotheses and express great confidence in their opinions even in the face of contradictory evidence or simply when guessing about an ambiguous matter.

Other factors that prevent scientists from recognizing the biases in their work are sociological and social psychological in character. For example, Lacy and Busch (1983) found that scientists reported communicating with colleagues outside their departments less than once a month, although they also reported that contacts with such colleagues were a major source of their research ideas. As a result, the authors point out, their communication takes on an "insular" pattern. Similarly, Mahoney (1985) notes that disciplinary structures, peer review processes, and tenure proceedings all tend to discourage the free exchange of novel and critical ideas essential to epistemic progress. Descriptions of how these matters can lead to problems in science abound. For examples, I point to Franks's (1981) description of the polywater episode and to Watson's (1968) description of DNA research.

The net effect of these social and psychological factors is that scientists are probably not sufficiently skeptical of their own work and that they probably do not receive sufficient criticism from diverse sources. Each factor is a two-edged sword! Both are reasons why scientists need criticism, and both are reasons why we can expect that scientists will resist criticism. If scientists are often too finite to recognize the multiple biases associated with their own work, if they too rarely seek and use serious criticism to help uncover those biases, and if disciplinary and other sociological structures often protect them from criticism, how can we be optimistic that critical multiplism will be adopted unless these factors are changed? Introducing science to critical multiplism may require changing the sociology and psychology of science. But we know that wholesale changes in social structures and personal beliefs and behaviors do not occur quickly or easily (Lindblom, 1977; Lindblom and Cohen, 1979; Shadish, 1984). This is probably as true of scientific change as it is of any other form of social change (Neimeyer and Shadish, 1987).

Adopting Critical Multiplism at the Individual Level. Felix Franks (1981) suggested that the pathologies in science that gave rise to the polywater controversy are at least partly a function of the way in which individual scientists think about their work. As he noted, "Michael Faraday once expressed the hope that fifty years after his death nothing he had ever written would still be considered true. Most scientists of today lack Faraday's modesty; they like to believe that everything they commit to paper will endure as truth forever. Genuine changes of heart or mind are therefore suspect" (Franks, 1981, p. 190).

Getting individual scientists to change this mind-set and adopt a critical multiplist perspective on their work may require instilling this attitude during initial training. Many scientific habits and ways of thinking are probably set during that time, and they may be more difficult to change later in a scientific career. If I were to design a course on research design for graduate students in psychology, for example, a reading outlining the tenets of critical multiplism and showing how it can be applied would figure prominently, although whether this reading should occur early or later in the course could be debated. Moreover, such courses often ask students to criticize the methodology of published studies. It should prove easy to add a critical multiplist component to such exercises first by selecting multiple studies that illustrate the topic being discussed, then by having students suggest new ways of doing the study and criticize any constant biases they found in the ones already done. Critical multiplism could also figure prominently in any courses discussing cognate ideas from the philosophy of science, such as logical positivism—assuming, of course, that novice scientists still take such courses.

Nevertheless, there are likely to be fundamental limits on the degree to which individual scientists can be critically multiplistic, if for no other reason than human finitude. For example, it is unrealistic to expect any single scientist to know how to conduct every possible statistical technique that might be relevant, from survival analysis to linear structural modeling via logistic regression. It is only slightly more realistic to hope that, with diligent effort that will probably have to extend well beyond graduate training, scientists can learn to recognize that there are relevant alternative options and that it might be appropriate to seek help in implementing them—so that they at least will know what survival analysis is, what it does, and when it can be relevant to their work, even if they cannot conduct such analysis themselves. Ultimately, individual scientists must fail even at this latter task, which is why the social component of critical multiplism is essential and also why it is wrong to think of critical multiplism as a state of mind of the individual scientist without a necessary social component.

In much the same way, even if one agrees in principle with the idea that scientists should be skeptical critics of their own and others' work, the demands of scientific practice can militate against the principle. Practice prefers the simple, the black-and-white, the easily implementable and codifiable course of action. A constant search for complexity could easily paralyze practice. Hence, there will inevitably be tension between critical multiplism and the demands of practice that the corollary technical guidelines outlined earlier can only partly remedy.

Finally, there may also be limitations on the degree to which scientists should be critically multiplistic, at least for some scientists performing some tasks in some stages of research. Mitroff's study of the Apollo moon scientists (Mitroff and Fitzgerald, 1977) gives most pause for concern here.

Mitroff found that the scientists whom their peers regarded as the most creative, influential, and visionary were the scientists who most stubbornly pursued their single favorite theory, even in the face of potentially serious conflicting evidence. Such pursuit is somewhat antithetical to critical multiplism, yet it is nonetheless justified both to test the limits of a theory (Lakatos, 1978) and to allow for the modicum of personal passion that is probably one essential ingredient of good science. As Popper (1972, p. 12) put it, "critical reasoning is better than passion, especially in matters touching on logic. But I am quite ready to admit that nothing will ever be achieved without a modicum of passion." Thus, some scientists may work best without being very self-critical. Yet, having successfully (and passionately) applied critical multiplism myself to the analysis of several theoretical and methodological problems (Houts, Cook, and Shadish, 1986; Shadish, Cook, and Houts, 1986), I am reluctant to endorse Mitroff's description of science as uniformly best for all good scientists. At a minimum, then, we need much more research on the social psychology of science if we are to understand when and why both passion and criticism should enter the research process (Gholson, Shadish, Neimeyer, and Houts, 1989; Shadish and Fuller, in press). I am therefore much more tentative about advocating critical multiplism at the individual level than I am at the social and institutional level.

Adopting Critical Multiplism at the Institutional Level. It seems more likely that the successful introduction of critical multiplism into science will have to come at the institutional and social level. This has long been Don Campbell's argument, and his recent work on the sociology of scientific validity bears witness to his continued creativity in suggesting practical ways in which science can be both critical and multiple. Perhaps the best example is his article on monitoring the scientific competence of the National Institute of Mental Health's (NIMH) Preventive Intervention Research Centers (Campbell, 1987). He makes a number of suggestions for improving scientific competence in these centers: Fund annual conferences on prevention research that would enable center staff to share current progress. Encourage problem overlap in research both within and between centers, and foster diversity of method among individual approaches to the same problem. Allow several scientists in the same center to submit independent grant applications. Split large studies into two or more parallel studies. Fund cross-validation research for the implementation of promising prevention interventions. Facilitate reanalysis and meta-analysis. Encourage grant final reports to include an appendix indicating how the investigator would redesign the study if he or she could do it over again. Legitimate and facilitate supplementary and dissenting-opinion research reports from research staff in forms ranging from simple dissenting footnotes to secondary analysis and publication by dissenting authors of alternative findings and interpretations. Finally, give preference

for funding to authors with established interdisciplinary publication records over those with narrower disciplinary focuses. Of course, not all these suggestions will prove effective or practical (Neimeyer and Shadish, 1987). But on the whole, too few theorists—including me—are thinking of ways that might make critical multiplism more widely used in science. That is one of the tasks on my own list of future articles. Undoubtedly, I will borrow heavily from the work of sociologists of science and from science policy analysts (Averch, 1985; Latour, 1987) in this kind of elaboration.

Practical Problems. Even if one were to accept all these points in principle, a few practical problems would make the use of critical multiplism difficult even for the truest of true believers. One is that a manuscript or a grant application developed in a critical multiplist mode is not likely to be well received. Critical multiplism is a tool for generating and criticizing options before you begin writing about those ideas. But once you have finished developing the ideas and you begin writing, the practical task is different—to weave as compelling a story as you can about the ideas on which you eventually settled. At that point, a complete list of the rejected options is distracting. One should no more lead the reader through the process that generated the good ideas than one should lead the reader through all the steps of every statistical analysis that one does of the data. In grant proposals and journal articles, we mostly show readers the results of our work, not the processes that we went through in getting to those results. It might be better if this were not so, but this is probably a fact of scientific writing.

Another worry raised by critical multiplism is that political or social reasons will make some novel options not very acceptable to colleagues, funding agencies, or society. For instance, one administrator in the National Institutes of Health told me that he would love to have grant proposals that looked at problems in truly novel ways but that, in his experience, review panels, especially those dominated by one discipline or specialty, tended to reject proposals for research that obviously departed from accepted paradigms. This is a good point. It takes a good deal of seasoned judgment and experience to know just exactly how novel one can afford to be and still get funded or published. Similarly, it is risky to submit a proposal that questions fundamental assumptions about the way in which things get done or that conflicts in important ways with the political and economic structures within which one works. The reason is not just the political risks incumbent in such proposals but also that the truly radical proposals would be extremely difficult to implement in public policy, no matter how effective they might prove to be. Inevitably, some of the options generated by a critical multiplist approach will be of this type, and they should be proposed only with great hesitation—particularly by younger, untenured, or otherwise insecure researchers. The best advice

here is to get input from an experienced scholar before submitting such a proposal—not just any scholar but one who has experience with review panels or publication outlets like the one to which you plan to submit.

Having said this, it is incumbent on me to note that many of the options generated through critical multiplism will not be antiestablishment. Rather, they will simply identify gaps in our knowledge that the granting agencies themselves may welcome having filled. For example, a colleague of mine has long been interested in research to prevent smoking. He started by trying interventions to help people quit. But there was always a hard core of smokers who would not quit or who soon relapsed if they did quit. On exploring their reasons, he discovered that they often complained that they gained weight after they quit, and research verified the complaint. He explored differences between males and females and found both that the weight gain was much more pronounced in females and that females had rarely been studied in smoking research. Hence, he proposed an array of projects exploring this problem in females. Was it antiestablishment? Quite the contrary. Agencies like the NIH have recognized the weaknesses of the monism in so much of our biomedical and health services research under which even the rats are white males. Such agencies were quite anxious to receive proposals on other populations. As these examples suggest, taking a critical multiplist stance often helps generate ideas that are neither antiestablishment or antiparadigm but rather that point to important gaps in our knowledge that many people want to see filled.

Finally, it is not possible to be fully critically multiplist in any given study. Time, resource, and skill constraints simply make it impossible to follow all options at once. Nor is it necessarily desirable to be fully critical multiplist in a series of closely related studies on a given problem. After all, one could not replicate a finding or fine-tune one's understanding of the results of a study if one changed the original conceptualization, design, execution, or analysis in any radical way. Rather, critical multiplism is probably a goal better suited for certain tasks and levels of effort than others: for programs of research rather than single studies; for the "invisible colleges" of researchers working competitively and cooperatively on the same problem whose members carve out a niche for themselves by identifying the gaps and problems in the relevant literature; for reconceptualizing an area in which progress has reached a standstill; for making sure that one's research is not getting into a conceptual or methodological rut; for generating important variations within one's program of research that are likely to make a real difference to the outcome; and, within single studies, for increasing the odds that one has not overlooked an important version of a question, an important measure, an important analysis, an important objection to what one has done, or any other of the reasonably inexpensive options that we could include in our studies with very little extra cost or effort.

Perhaps the point to remember in all this is the value of criticism, whether it be critical multiplism in one's program of research or critical monism in a single study. In all cases, it is always better to know one's strengths and weaknesses. But it is easier to be multiplist than it is to be critical. Good criticism is a precious commodity, difficult to find the energy to do and difficult to hear from others. Good criticism does not just come up with alternatives but with alternatives that make an important difference for the results.

Conclusion

A colleague prominently displays the following quotation on his office door:

When someone is honestly 55 percent right, that's very good and there's no use wrangling. And if someone is 60 percent right, it's wonderful, it's great luck, and let him thank God. But what's to be said about 75 percent right? Wise people say this is suspicious. Well, and what about 100 percent right? Whoever says he's 100 percent right is a fanatic, a thug, and the worst kind of rascal.

The old Jew of Galicia to whom this remark is attributed has captured the kind of skepticism advocated in critical multiplism. Probably no scientist and no scientific study are ever 100 percent right. We can only hope that, if we do our job properly and with sufficient critical perspective, we may be slightly more often right than wrong—55 percent right, so to speak. But we should always keep in mind how close we are to being more often wrong than right—even a slip to 45 percent right would mean that we would be mostly wrong. In social science, we often work in this small margin. We need every tool at our disposal to help us to produce results that in the long run fall on the right side of the distribution.

References

- Ashmore, M. *The Reflexive Thesis: Wrighting the Sociology of Scientific Knowledge*. Chicago: University of Chicago Press, 1989.
- Averch, H. A. *A Strategic Analysis of Science and Technology Policy*. Baltimore, Md.: Johns Hopkins University Press, 1985.
- Bechtel, W. *Philosophy of Science: An Overview for Cognitive Science*. Hillsdale, N.J.: Erlbaum, 1988.
- Bem, D. J. "Constructing a Theory of the Triple Typology: Some (Second) Thoughts on Nomothetic and Idiographic Approaches to Personality." *Journal of Personality*, 1983, 51, 566–577.
- Bem, D. J., and Allen, A. "On Predicting Some of the People Some of the Time: The Search for Cross-Situational Consistencies in Behavior." *Psychological Review*, 1974, 81, 506–520.

- Boruch, R. F. "On Common Contentions About Randomized Field Experiments." In R. F. Boruch and H. W. Reicken (eds.), *Experimental Testing of Public Policy: The Proceedings of the 1974 Social Science Research Council Conference on Social Experiments*. Boulder, Colo.: Westview Press, 1975.
- Boruch, R. F., McSweeney, A. J., and Soderstrom, E. J. "Randomized Field Experiments for Program Planning, Development, and Evaluation: An Illustrative Bibliography." *Evaluation Quarterly*, 1978, 2, 655-695.
- Bradburn, N. M., Sudman, S., Blair, E., Locander, W., Miles, C., Singer, E., and Stocking, C. *Improving Interview Methods and Questionnaire Design*. San Francisco: Jossey-Bass, 1979.
- Bradley, J. V. "Overconfidence in Ignorant Experts." *Bulletin of the Psychonomic Society*, 1981, 17, 82-84.
- Buss, D. M., and Craik, K. H. "The Act Frequency Approach to Personality." *Psychological Review*, 1983, 90, 105-126.
- Buss, D. M., and Craik, K. H. "Acts, Dispositions, and Personality." In B. A. Maher and W. B. Maher (eds.), *Progress in Experimental Personality Research*. Vol. 13. New York: Academic Press, 1984.
- Campbell, D. T. "Evolutionary Epistemology." In P. A. Schilpp (ed.), *The Philosophy of Karl Popper*. La Salle, Ill.: Open Court, 1974.
- Campbell, D. T. "Guidelines for Monitoring the Scientific Competence of Preventive Intervention Research Centers: An Exercise in the Sociology of Scientific Validity." *Knowledge: Creation, Diffusion, Utilization*, 1987, 8, 389-430.
- Campbell, D. T., and Fiske, D. W. "Convergent and Discriminant Validity by the Multitrait-Multimethod Matrix." *Psychological Bulletin*, 1959, 56, 81-105.
- Campbell, D. T., and Stanley, J. C. *Experimental and Quasi-experimental Designs for Research*. Skokie, Ill.: Rand McNally, 1963.
- Chen, H., and Rossi, P. H. "The Theory-Driven Approach to Validity." *Evaluation and Program Planning*, 1987, 10, 95-103.
- Cicirelli, V. G., and Associates. *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development. A Report to the Office of Economic Opportunity*. Athens: Ohio University and Westinghouse Learning Corporation, 1969.
- Cook, T. D. "Postpositivist Critical Multiplism." In L. Shotland and M. M. Mark (eds.), *Social Science and Social Policy*. Newbury Park, Calif.: Sage, 1985.
- Cook, T. D. "The Generalization of Causal Connections: Theories in Search of Clear Practice." In L. Sechrest, E. Perrin, and J. Bunker (eds.), *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*. Publication No. (PHS) 90-3435. Rockville, Md.: National Center for Health Services Research and Health Care Technology Assessment, Public Health Service, U.S. Department of Health and Human Services, 1990.
- Cook, T. D., and Campbell, D. T. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Skokie, Ill.: Rand McNally, 1979.
- Cook, T. D., and Reichardt, C. S. (eds.). *Qualitative and Quantitative Methods in Evaluation Research*. Newbury Park, Calif.: Sage, 1979.
- Coombs, C. *A Theory of Data*. New York: Wiley, 1964.
- Cronbach, L. J. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass, 1982.
- Cronbach, L. J., Gleser, G., Nanda, H., and Rajaratnam, N. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley, 1972.
- Davies, P. C. W., and Brown, J. R. *The Ghost in the Atom*. Cambridge, England: Cambridge University Press, 1986.
- Dawes, R. M. "Measurement Models for Rating and Comparing Risks: The Context of AIDS." In L. Sechrest, H. Freeman, and A. Mulley (eds.), *Health Services Research Methodology: A Focus on AIDS*. Publication No. (PHS) 89-3439. Rockville, Md.: National Center for Health Services Research and Health Care Technology Assessment, Public Health Service,

- U.S. Department of Health and Human Services, 1989.
- D'Espagnat, B. *In Search of Reality*. New York: Springer-Verlag, 1983.
- Dunn, W. N. "Reforms as Arguments." *Knowledge: Creation, Diffusion, Utilization*, 1982, 3, 293–326.
- Epstein, S. "The Stability of Behavior: On Predicting Most of the People Much of the Time." *Journal of Personality and Social Psychology*, 1979, 37, 1097–1126.
- Faust, D. *The Limits of Scientific Reasoning*. Minneapolis: University of Minnesota Press, 1984.
- Feyerabend, P. *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: Verso, 1975.
- Fisher, R. A. *Statistical Methods for Research Workers*. London: Oliver & Boyd, 1925.
- Franks, F. *Polywater*. Cambridge, Mass.: MIT Press, 1981.
- Garfield, E. "Is Citation Analysis a Legitimate Evaluation Tool?" *Scientometrics*, 1979, 1, 359–375.
- Gholson, B., and Barker, P. "Kuhn, Lakatos, and Laudan: Applications in the History of Physics and Psychology." *American Psychologist*, 1985, 40, 755–769.
- Gholson, B. G., Shadish, W. R., Neimeyer, R. A., and Houts, A. C. (eds.). *Psychology of Science: Contributions to Metascience*. Cambridge, England: Cambridge University Press, 1989.
- Glass, G. V., McGaw, B., and Smith, M. L. *Meta-analysis in Social Research*. Newbury Park, Calif.: Sage, 1981.
- Groopman, J. E., Salahuddin, S. Z., Sarngadharan, M. G., Markham, P. D., Gonda, M., Sliski, A., and Gallo, R. C. "HTLV-III in Saliva of People with AIDS-Related Complex and Healthy Homosexual Men at Risk for AIDS." *Science*, 1984, 226, 447–449.
- Haveman, R. H. "Social Experimentation and Social Experimentation." *Journal of Human Resources*, 1985, 21, 586–605.
- Hedges, L. V., and Olkin, I. *Statistical Methods for Meta-analysis*. Orlando, Fla.: Academic Press, 1985.
- Ho, D. D., Byington, R. E., Schooley, R. T., Flynn, T., Rota, T. R., and Hirsch, M. S. "Infrequency of Isolation of HTLV-III Virus from Saliva in AIDS." *New England Journal of Medicine*, 1985, 313, 1606.
- Houts, A. C., Cook, T. D., and Shadish, W. R. "The Person-Situation Debate: A Critical Multiplist Perspective." *Journal of Personality*, 1986, 54, 101–154.
- Keller, E. F. *A Feeling for the Organism*. San Francisco: Freeman, 1983.
- Kinsey, A. C., Pomeroy, W. B., and Martin, C. E. *Sexual Behavior in the Human Male*. Philadelphia: Saunders, 1948.
- Kuhn, T. S. *The Structure of Scientific Revolutions*. (2nd ed.) Chicago: University of Chicago Press, 1970.
- Lacy, W. B., and Busch, L. "Informal Scientific Communication in the Agricultural Sciences." *Information Processing and Management*, 1983, 19, 193–202.
- Lagakos, S. "Statistical Analysis of Survival Data, with Applications to the AIDS Epidemic." In L. Sechrest, H. Freeman, and A. Mulley (eds.), *Health Services Research Methodology: A Focus on AIDS*. Publication No. (PHS) 89-3439. Rockville, Md.: National Center for Health Services Research and Health Care Technology Assessment, Public Health Service, U.S. Department of Health and Human Services, 1989.
- Lakatos, I. *The Methodology of Scientific Research Programmes: Philosophical Papers*. Vol. 1. Cambridge, England: Cambridge University Press, 1978.
- LaLonde, R. J. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review*, 1986, 76, 604–620.
- Latour, B. *Science in Action*. Cambridge, Mass.: Harvard University Press, 1987.
- Latour, B., and Woolgar, S. *Laboratory Life: The Social Construction of Scientific Facts*. London: Sage, 1979.

- Laudan, L. *Progress and Its Problems: Towards a Theory of Scientific Growth*. Berkeley: University of California Press, 1977.
- Lewin, R. "A Naturalist of the Genome." *Science*, 1983, 222, 402–405.
- Lindblom, C. E. *Politics and Markets: The World's Political-Economic Systems*. New York: Basic Books, 1977.
- Lindblom, C. E., and Cohen, D. K. *Usable Knowledge: Social Science and Social Problem Solving*. New Haven, Conn.: Yale University Press, 1979.
- Lord, F. M., and Novick, M. R. *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley, 1968.
- Lowry, O. H., Rosenbrough, N. J., Farr, A. L., and Randall, R. J. "Protein Measurement with the Folin Phenol Reagent." *Journal of Biological Chemistry*, 1951, 193, 265–275.
- Magidson, J. "Toward a Causal-Model Approach for Adjusting for Preexisting Differences in the Nonequivalent Control Group Situation: A General Alternative to ANCOVA." *Evaluation Quarterly*, 1977, 1, 399–420.
- Mahoney, M. J. *Psychologist as Subject: The Psychological Imperative*. Cambridge, Mass.: Ballinger, 1976.
- Mahoney, M. J. "Open Exchange and Epistemic Progress." *American Psychologist*, 1985, 40, 29–39.
- Mark, M. M., and Shotland, R. L. "Toward More Useful Social Science." In R. L. Shotland and M. M. Mark (eds.), *Social Science and Social Policy*. Newbury Park, Calif.: Sage, 1985.
- Meehl, P. E. "What Social Scientists Don't Understand." In D. W. Fiske and R. A. Shweder (eds.), *Metatheory in Social Science*. Chicago: University of Chicago Press, 1986.
- Mischel, W. "On the Interface of Cognition and Personality: Beyond the Person-Situation Debate." *American Psychologist*, 1979, 34, 740–754.
- Mischel, W., and Peake, P. K. "Beyond Deja Vu in Search for Cross-Situational Consistency." *Psychological Review*, 1982, 89, 730–755.
- Mitroff, I. I., and Fitzgerald, I. "On the Psychology of the Apollo Moon Scientists: A Chapter in the Psychology of Science." *Human Relations*, 1977, 130, 657–674.
- Naisbitt, J. *Megatrends: Ten New Directions Transforming Our Lives*. New York: Warner Books, 1984.
- National Center for Health Statistics. "AIDS Knowledge and Attitudes for October 1987: Provisional Data from the National Health Interview Survey." In *Advance Data from Vital and Health Statistics*, no. 150. Publication No. (PHS) 88-1250. Hyattsville, Md.: Public Health Service, U.S. Department of Health and Human Services, 1988.
- Neimeyer, R. A., and Shadish, W. R. "Optimizing Scientific Validity: Toward an Interdisciplinary Science Studies." *Knowledge: Creation, Diffusion, Utilization*, 1987, 8, 463–485.
- Nunnally, J. C. *Psychometric Theory*. (2nd ed.) New York: McGraw-Hill, 1978.
- Popper, K. R. *Objective Knowledge: An Evolutionary Approach*. Oxford, England: Oxford University Press, 1972.
- Rae, A. *Quantum Physics: Illusion or Reality?* Cambridge, England: Cambridge University Press, 1986.
- Rein, M., and Schon, D. A. "Problem Setting in Policy Research." In C. H. Weiss (ed.), *Using Social Research in Public Policy-Making*. Lexington, Mass.: Lexington Books, 1977.
- Rindskopf, D. M. "Structural Equation Models in Analysis of Nonexperimental Data." In R. F. Boruch, P. M. Wortman, and D. S. Cordray (eds.), *Reanalyzing Program Evaluations*. San Francisco: Jossey-Bass, 1981.
- Rossi, P. H., Berk, R. A., and Lenihan, K. J. *Money, Work, and Crime: Some Experimental Evidence*. New York: Academic Press, 1980.
- Rossi, P. H., Wright, J. D., and Anderson, A. (eds.). *Handbook of Survey Research*. New York: Academic Press, 1983.
- Schreeder, M. T., Thompson, S. E., Hadler, S. C., Berquist, K. R., Zaidi, A., Maynard, J. E., Judson, N., Broff, E. H., Nylund, T., Moore, J. N., Gardner, P., Doto, I. L., and Reynolds,

- G. "Hepatitis B in Homosexual Men: Prevalence of Infection and Factors Related to Transmission." *Journal of Infectious Diseases*, 1982, 146, 7-15.
- Shadish, W. R. "Policy Research: Lessons from the Implementation of Deinstitutionalization." *American Psychologist*, 1984, 39, 725-738.
- Shadish, W. R. "Planned Critical Multiplism: Some Elaborations." *Behavioral Assessment*, 1986, 8, 75-103.
- Shadish, W. R., Cook, T. D., and Houts, A. C. "Quasi-experimentation in a Critical Multiplist Mode." In W.M.K. Trochim (ed.), *Advances in Quasi-experimental Design and Analysis*. New Directions for Program Evaluation, no. 81. San Francisco: Jossey-Bass, 1986.
- Shadish, W. R., Cook, T. D., and Leviton, L. C. *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, Calif.: Sage, 1991.
- Shadish, W. R., Doherty, M., and Montgomery, L. M. "How Many Studies Are in the File Drawer? An Estimate from the Family/Marital Psychotherapy Literature." *Clinical Psychology Review*, 1989, 9, 589-603.
- Shadish, W. R., and Fuller, S. (eds.). *Social Psychology of Science: The Psychological Turn*. New York: Guilford, in press.
- Shadish, W. R., and Reis, J. "A Review of the Effectiveness of Programs to Improve Pregnancy Outcome." *Evaluation Review*, 1984, 8, 747-776.
- Sharpe, T. R., and Wetherbee, H. *Final Report: Evaluation of the Improved Pregnancy Outcome Program*. Tupelo, Miss.: Three Rivers District Health Department, Mississippi State Board of Health, 1980.
- Smith, M. L., Glass, G. V., and Miller, T. I. *The Benefits of Psychotherapy*. Baltimore, Md.: Johns Hopkins University Press, 1980.
- Thompson, M. S., and Meyer, H. J. "The Costs of AIDS: Alternative Methodological Approaches." In L. Sechrest, H. Freeman, and A. Mulley (eds.), *Health Services Research Methodology: A Focus on AIDS*. Publication No. (PHS) 89-3439. Rockville, Md.: National Center for Health Services Research and Health Care Technology Assessment, Public Health Service, U.S. Department of Health and Human Services, 1989.
- Tukey, J. W. *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley, 1977.
- Turner, C. F. *Public Perceptions and Behavior in Response to the AIDS Epidemic: Present Knowledge and Future Needs*. Unpublished manuscript, CBASSE Committee on AIDS Research, National Research Council, Washington, D.C., 1988.
- Turner, C. F., Fay, R. E., and Widdus, R. *Monitoring the Spread of HIV Infection*. Unpublished manuscript, CBASSE Committee on AIDS Research, National Research Council, Washington, D.C., 1988.
- Valdiserri, R. O., Lyter, D. W., Kingsley, L. A., Leviton, L. C., Schofield, J. W., Huggins, J., Ho, M., and Rinaldo, C. R. "The Effect of Group Education on Improving Attitudes About AIDS Risk Reduction." *New York State Journal of Medicine*, May 1987, pp. 272-278.
- Valdiserri, R. O., Lyter, D. W., Leviton, L. C., Stoner, K., and Silvestre, A. *Applying the Criteria for the Development of Health Promotion and Education Programs to AIDS Risk Reduction Programs for Gay Men*. Unpublished manuscript, Graduate School of Public Health, University of Pittsburgh, 1988.
- Watson, J. D. *The Double Helix*. New York: Atheneum, 1968.
- Woolgar, S. (ed.). *Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge*. London: Sage, 1988.
- Zwick, W. R., and Velicer, W. F. "Comparison of Five Rules for Determining the Number of Components to Retain." *Psychological Bulletin*, 1986, 99, 432-442.

WILLIAM R. SHADISH is professor of psychology at Memphis State University. He is author (with Thomas D. Cook and Laura C. Leviton) of Foundations of Program Evaluation and current editor of New Directions for Program Evaluation.