Another Look at Meehl, Lakatos, and the Scientific Practices of Psychologists

Reuven Dar  University of Wisconsin—Madison

ABSTRACT: This article examines current research methodology in psychology in the context of Serlin and Lapsley’s response to Meehl’s critiques of the scientific practices of psychologists. The argument is made that Serlin and Lapsley’s appeal to Lakatos’s philosophy of science to defend the rationality of null hypothesis tests and related practices misrepresents that philosophy. It is demonstrated that Lakatos in fact considered psychology an extremely poor science lacking true research programs, an opinion very much in line with Meehl’s critique. The present essay speculates on the reasons for Lakatos’s negative opinion and reexamines the role of null hypothesis tests in relation to the quality of theories in psychology. It is concluded that null hypothesis tests are destructive to theory building and directly related to Meehl’s observation of slow progress in soft psychology.

In a recent article Serlin and Lapsley (1985) reviewed Meehl’s (1967, 1978) critique of the scientific practices of psychologists. According to Serlin and Lapsley, some of Meehl’s criticism is not justified; specifically, they suggested that from the perspective of the philosophy of science developed by Imre Lakatos (1978a), most of the actual practice of scientific psychology can be rationally defended.

The present article is an attempt to further examine Meehl’s critique and Serlin and Lapsley’s defense of scientific psychology, particularly of the use of null hypothesis tests. In the following pages I argue that although Serlin and Lapsley’s examination of the rationality of research methodology in psychology is important and timely, the conclusions they drew in regard to this examination are a bit too positive. I take issue with some of Serlin and Lapsley’s points, particularly with their interpretation of Lakatos’s philosophy as it applies to scientific psychology. I also try to offer an additional perspective on the issues raised by Meehl and particularly on the implications of null hypothesis testing for the research in “soft” psychology. First, however, it is necessary to review briefly the critique offered by Meehl, on the one hand, and Serlin and Lapsley’s response, focusing on their appeal to Lakatos’s philosophy, on the other.

Meehl’s Critique and Serlin and Lapsley’s Response

Meehl’s original critique of theory testing in psychology (Meehl, 1967) was focused on what Serlin and Lapsley (1985) called “the asymmetry argument (p. 74).” Meehl observed that whereas physicists derive and test point predictions, psychologists do the opposite: They test the theory by attempting to reject a point-null hypothesis, which is the logical complement of the theoretical prediction. This, argued Meehl, leads to the paradox that in psychology, increased measurement precision makes it easier for the theory to pass the test, whereas in physics it increases the severity of the test, as presumably it should. The tests to which theories in psychology are subjected, wrote Meehl, are so easy that passing them is “only an extremely weak corroboration of any substantive theory” (Meehl, 1967, p. 111). This problem is compounded, according to Meehl, by the practice of evaluating a theory by “counting noses,” that is, comparing the number of experimental ‘confirmations’ to the number of ‘refutations’ of the theory under test (p. 112). This, said Meehl, is scientifically preposterous, because, as deductive logic dictates, refutations are much more critical to the validity of a theory than confirmations. An additional problem is that the typical theory in psychology is only loosely associated, through its “auxiliary hypotheses,” to its operations in the experiment. This, according to Meehl, means that the results of the experimental test, which is absurdly easy to begin with, have no real impact on the substantive theory itself. It also leads to excessive ad hoc theorizing in which the auxiliary hypotheses are challenged and tested repeatedly while the substantive theory is left behind, never subjected to a risk of refutation.

More recently, Meehl (1978) criticized the slow, noncumulative scientific progress in “soft” psychology. Elaborating on his 1967 arguments, Meehl tied this slow progress to faulty scientific methodology and especially to the reliance on null hypothesis tests.

Serlin and Lapsley’s (1985) article has two distinct components. On the one hand, the authors agreed with Meehl on the need to increase the severity of the tests to which psychologists subject their theories; to that end they proposed the methodology of the “good-enough principle,” which is a most sophisticated solution to this problem. On the other hand, they rejected other aspects of Meehl’s critique by appealing to Lakatos’s (1978a) views on the history and philosophy of science; this latter component is the focus of the present essay.

Serlin and Lapsley (1985) responded to Meehl’s critique with an attempt at “reconstructing the actual practice of psychologists into a logically defensible form” (p. 73). They asserted that null hypothesis testing, in principle, is necessary and should be a part of the methodology of the hard as well as the soft sciences. They also suggested
that much of Meehl's criticism reflects his failure to apply Lakatos's "sophisticated" (as opposed to Popper's "naive") methodological falsificationism (Lakatos, 1978a) to psychology. Specifically, Serlin and Lapsley maintained that Meehl's complaints against such practices as "nose counting," ad hoc explanations, and repeated appeals to auxiliary hypotheses, as well as his observation of seemingly slow progress in "soft" psychology, can all be accommodated in Lakatos's framework.

**Lakatos's Actual Views on the Social Sciences**

Serlin and Lapsley did not purport to represent the late Lakatos's actual views on the scientific practices in psychology. Still, they did appeal to Lakatos's philosophy to resolve the issues raised by Meehl (1967, 1978), which seems to imply that Lakatos held a fairly favorable opinion in regard to the scientific practice of psychologists. Such, however, is not the case. Lakatos, in fact, was clearly influenced by Meehl's 1967 critique, which appeared in a philosophy journal, and the following quote from Lakatos preceded Meehl's 1978 critique, which was published after Lakatos's death. In his paper entitled "Falsification and the Methodology of Scientific Research Programs," Lakatos (1978a) wrote,

> This requirement of continuous growth . . . hits patched-up, unimaginative series of pedestrian "empirical" adjustments which are so frequent, for instance, in modern social psychology. Such adjustments may, with the help of so-called "statistical techniques," make some "novel" predictions and may even conjure up some irrelevant grains of truth in them. But this theorizing has no unifying idea, no heuristic power, no continuity. They do not add up to a genuine research programme and are, on the whole, worthless. (p. 88)

In a footnote Lakatos (1978a) referred more specifically to the use of those "statistical techniques":

> After reading Meehl (1967) and Lykken (1968) one wonders whether the function of statistical techniques in the social sciences is not primarily to provide a machinery for producing phoney corroborations and thereby a semblance of "scientific progress" where, in fact, there is nothing but an increase in pseudo-intellectual garbage. (p. 88)

Lakatos, then, did not intend his philosophy of science to be used to defend the rationality of the practice of social scientists, as it was used by Serlin and Lapsley—quite the contrary, in fact, as the following words testify:

> "Thus the methodology of research programs might help us in devising laws for stemming this intellectual pollution which may destroy our cultural environment even earlier than industrial and traffic pollution destroys our physical environment" (Lakatos, 1978a, p. 89).

Clearly, Lakatos's expressed hostility toward the social sciences cannot be taken as evidence that he thoroughly examined the practice of social scientists in light of his philosophy and determined that it was irrational and nonscientific; Lakatos's view may well have been influenced by other, personal factors. It is quite intriguing, however, to observe that the implications for the social sciences that Serlin and Lapsley drew from Lakatos's philosophy are so discrepant from Lakatos's own views on the matter and that, in fact, as far as we can tell, Lakatos was in complete agreement with Meehl's critique. Before discussing possible reasons for this discrepancy, a brief description of Lakatos's philosophy of science may be helpful; specific points in his scheme will be highlighted later as they become relevant to the discussion (for a more complete picture, see Lakatos, 1978a).

Lakatos's position takes off from the philosophy of science of his teacher, Sir Karl Popper. Lakatos's "sophisticated methodological falsificationism" departs from Popper's conception in focusing on "research programs" rather than on isolated theories and, more important, in rejecting Popper's notion of progress by strict refutation. According to Lakatos, an experimental "refutation," or a demonstration of inconsistency, does not eliminate a research program. In fact, the "hard core" of a research program, being protected by the "negative heuristic," is never subjected to the danger of empirical refutation. It is often rational, according to Lakatos, to continue and develop a program, guided by its "positive heuristic," in the face of such "refutations." A research program should be abandoned only if it is clearly "degenerating" and if rival programs demonstrate their superiority. In general, then, Lakatos took a more liberal approach to the philosophy of science than the one represented by Popper, and by doing so he accounted better for the actual tenacity of theories in the face of so-called empirical refutations.

With this synopsis in mind, let us now turn back to the discrepancy question introduced earlier: Why was Lakatos in disagreement with the favorable implications that Serlin and Lapsley drew from his philosophy in regard to the scientific practices of psychologists? This discrepancy is especially puzzling when we note that some of Meehl's criticism, as Serlin and Lapsley (1985) rightly pointed out, does seem to reflect a Popperian "naive" methodological falsificationism (Lakatos, 1978a). This is evident in Meehl's (1978) dismay at the persistence of theories in soft psychology in the face of experimental "refutations" (i.e., failures to reject the null hypothesis). The view that one takes the whole theory to test every time one performs an experiment was rejected by Lakatos, as mentioned previously. Thus Meehl's complaint that in psychology theories are not abandoned when the numerical results are not as forecasted can easily be accommodated in the Lakatosian framework. The purpose of the present essay, however, is to argue that Meehl's objections to the practices involved in null hypothesis testing pose a more severe challenge that cannot be resolved, as Serlin and Lapsley (1985) claimed, by an appeal to Lakatos's reconstruction of science. In the following para-
graphs I propose that Lakatos's comments about psychology are quite consistent with his philosophy, whereas Serlin and Lapsley's (1985) more favorable interpretation is not. I suggest that Lakatos saw problems in psychology that overshadow the specific practices that Serlin and Lapsley (1985) defended, by referring to Lakatos's philosophy, in their article. I also address the role of null hypothesis testing in contributing to both Meehl's and Lakatos's negative evaluation of scientific psychology. Specifically, I suggest that the weakness of theory testing in psychology cannot be dissociated from other scientific practices and that the role that null hypothesis testing has in slowing progress in psychology (Meehl, 1978) is due to its destructive effect on theory building. To support these contentions, I now turn to a more detailed review of Meehl's criticism and of Serlin and Lapsley's responses.

"Nose Counting" and the Severity of Theory Testing

As mentioned earlier, one of Meehl's central complaints in both of his essays is that psychology researchers, in a typical review article, commonly count the number of favorable and unfavorable experimental results to see how well their theory has fared in the experimental test. This "nose count," Meehl asserted, is absurd, because adverse results, "seen properly, do the theory far more damage than the favorable ones do it good" (Meehl, 1978, p. 822). According to Serlin and Lapsley, however, "nose counting" is quite rational in the Lakatosian scheme of things. They reminded us that in Lakatos's view a theory is never refuted by a single experiment and that disagreements with "facts" are treated as temporary anomalies. Thus it is rational to emphasize the empirical "successes" of the theory and not to get unduly nervous about its temporary "failures." I suggest, however, that this is too liberal an interpretation of Lakatos's ideas. Specifically, I believe that the "nose counting" issue cannot be separated from the issue of the severity of theory testing in psychology as compared, to use the example chosen by Meehl (1967), to physics and that Lakatos would have accepted "nose counting" in physics, but not in psychology. When a physicist conducts 10 experiments in which he or she predicts a specific value or function and misses slightly in 3 of them, then indeed his or her theory is not in bad shape. But it seems to me that when a psychologist fails to obtain results significantly different from zero (assuming a reasonably powerful design), one is forced to agree with Meehl that the relevant theory is indeed not in good shape. This is, of course, because the test of the psychologist's theory was so much easier to begin with, with such high initial probability for passing (see Meehl, 1967) that failing the test is much more significant for the validity of the theory than passing it. If and when psychologists start using Serlin and Lapsley's methods to create more severe tests for their theories, then it will become rational to treat successful predictions as seriously as failures; but in the meantime, Meehl's critique on this point seems compelling.

The general argument that can be made here is that Lakatos's philosophy cannot be applied to psychology without modifications, as Serlin and Lapsley (1985) attempted to do. Lakatos himself was a mathematician, and his rational reconstruction of science was founded on the history of the exact, or "hard," sciences, especially physics. In these sciences, the demand for powerful theories that make specific prediction, as well as for severe tests of theories, is an established tradition. It is my contention that when Lakatos's philosophy gives a particular science the license to be liberal in dealing with anomalies and in allowing theories to continue to exist in the face of experimental refutations, it assumes that that science is demanding in terms of theory building and testing. I suggest that Lakatos's expressed disdain for psychology and other social sciences, which is at odds with Serlin and Lapsley's interpretation of his philosophy, stems from his belief that these sciences do not make much demands on theories—a belief that Lakatos, as shown earlier, expressed clearly. Let us now consider the implications of this argument for another issue: that of ad hoc explanations in psychology.

Ad Hoc Explanations in Psychology

According to Meehl (1967), there is "a fairly widespread tendency to report experimental findings with a liberal use of ad hoc explanations for those that didn't 'pan out'" (p. 114), thus avoiding the implication of the experimental failure for the theory. In discussing the issue of ad hoc explanations it would be useful to distinguish, following Lakatos (1978a), between two types of ad hoc theorizing. The first one refers to the common usage of this term, involving hypotheses that are generated solely by the data, isolated from any theory or research program. The second type of ad hoc theorizing refers to nonbold theories that predict no novel facts and is more relevant to Meehl's complaint against the excessive challenging of auxiliary hypotheses in psychology, to be discussed in the next section.

Although Serlin and Lapsley (1985) agreed that the first type of ad hoc theorizing should be condemned, they denied the implication that "such 'naive guessing' constitutes the majority of social science research" or that "there are no research programs in psychological research" (p. 78). Notably, Serlin and Lapsley were not denying that such ad hoc explanations are at least fairly endemic to psychological research, and I speculate that they would also agree that such explanations are less common in the more exact sciences.

Granted that, what may be the reasons for the relative prevalence of this type of ad hoc hypothesis in psychology? I would like to suggest that it is again related to the fact that our theories generate very weak predictions. I dare say that, more often than not, either success or failure in a psychological experiment can be accounted for by a variety of improvised, ad hoc explanations. Let us examine this situation more closely. When a researcher in psychology predicts a difference between two groups without committing himself or herself to the size of that difference, it seems that one of two consequences is in-
evitable: (a) If a difference is indeed obtained (i.e., a difference significantly greater than zero), there will be numerous alternative explanations for it, probably just as reasonable as the researcher’s pet theory. This is because it is easy to imagine numerous states of nature that would lead to a nonzero difference between two groups. (b) If the predicted difference is not obtained (i.e., is not significantly greater than zero), then there will be several reasonable explanations for that, too, because many factors can be imagined that would diminish a small difference between two groups.

I personally feel that the state of affairs illustrated above is one important reason for the fact that so many theories in “soft” psychology do not endure (see Meehl, 1978). How can a theory be compelling if it is not, at least in some respects, unique in its ability to account for some experimental results? When it is so easy to come up with competing explanations for successful predictions and with ad hoc explanations for failing ones, why should the theory be taken seriously?

In physics, on the other hand, theories are tighter and lead to precise predictions. As a consequence, (a) if the numerical result is as predicted (that is, close enough to the predicted point value or curve), it will be very difficult, in contrast to the situation in psychology, to offer a reasonable alternative theory for that. This is because it is difficult to imagine alternative states of nature that will lead to the exact same curve or numerical result. (b) If the experimental result is not as predicted, some serious revision of the theory would be required. This is because a tight theory simply does not allow for significant (I do not mean “statistically significant”) discrepancies from predicted outcome. This point is elaborated on in the next section, in which I also attempt to speculate further on Lakatos’s position in regard to the two kinds of ad hoc explanations.

Auxiliary Hypotheses in Psychology

In psychology, Meehl observed, a typical experiment “involves use of complex and rather dubious auxiliary assumptions, which are required to mediate the original prediction and are therefore readily available as (genuinely) plausible ‘outs’ when the prediction fails” (Meehl, 1967, p. 114). Indeed, observed Meehl, it is typical in psychology to have a series of experiments in which each is designed to solve an “anomaly” found in the previous experiment.

In this fashion a zealous and clever investigator can slowly wend his way through a tenuous nomological network, performing a long series of related experiments which appear to the uncritical reader as a fine example of an “integrated research program,” without ever refuting or corroborating so much as a single strand of the network. (Meehl, 1967, p. 114, emphasis in original)

In defending the rationality of this second type of ad hoc explanation, Serlin and Lapsley reminded us that, according to Lakatos, the auxiliary hypotheses are actually the most appropriate targets for reevaluation if the experiment fails. In fact, according to Serlin and Lapsley (1985), “the rational procedure is to do what Meehl found preposterous, to examine empirical discrepancies by thoroughly testing the *ceteris paribus* clause” (p. 77). But although Serlin and Lapsley implied that in this respect the hard and the soft sciences are on an equal footing, I suggest that they are not. To convince oneself of that, one need only turn to Meehl’s entertaining (though quite realistic) example of a social psychological experiment designed to investigate the effect of social fear on visual perception (Meehl, 1978). In this experiment, as is quite typical in “soft psychology,” both the manipulation chosen to represent the independent variable and the instrument intended to assess the dependent variable can be easily challenged. And although Meehl recognized that this problem is not qualitatively unique in the social sciences, he insisted (and I agree) that it is quantitatively more severe for us. Meehl mentioned two reasons for that: First, because of the “looseness of the nomological network” in psychology it is quite difficult to conduct independent tests of auxiliary theories. Second (and more relevant to the present article), in the hard sciences there seems to be a “more intimate” connection between the substantive theory and the auxiliary, “sometimes one of contributing to derivability” (Meehl, 1978, p. 819).

Thus, as Meehl pointed out, in the hard sciences the measurement tools are much more integrated into the substantive theory. Moreover, they are often considered widely accepted scientific knowledge, having acquired the status that Lakatos termed background knowledge, and thus cannot be used as an easy way out of an experimental failure. This is clearly not the case in psychology; in fact, I cannot think of a single important construct in psychology for which there exists only a single agreed-on measurement tool. So although an anomaly encountered in a psychological experiment can be dealt with superficially by resorting to those “dubious auxiliary assumptions,” it is not as simple in the hard sciences. An off-handed resort to the “validity of the questionnaire” or the social psychologists’ favorite “demand characteristics” is not an available escape for the physicist.

The issue, then, is not whether the auxiliary hypotheses should be examined. Indeed, according to Lakatos (1978a), and as noted by Serlin and Lapsley, usually only the auxiliary hypotheses are being challenged by the scientist, whereas the “hard core” of the theory is protected from the *modus tollens*. The emphasis here, however, is on the *ease* with which it can and is being done in psychology. This ease seems to essentially diminish the distinction between the two kinds of ad hoc explanations: Because we are not committed to much widely accepted knowledge in “soft” psychology, our ad hoc challenges to auxiliary hypotheses have so little reference to theory that they become indistinguishable from the first (and more clearly condemnable) type of ad hoc explanation.

How does Lakatos’s philosophy figure in this issue? As shown earlier, Lakatos (1978a) did not think highly of the use of ad hoc explanations in psychology, describing psychological theories as “patched up” and being merely “series of pedestrian ‘empirical’ adjustments” (p. 88). It
seems clear that when Lakatos allowed for ad hoc explanations as a rational scientific behavior, he did not have psychology in mind. The reason for that, I suggest, is that Lakatos demanded specifically that the auxiliary theories be outlined in advance in the positive heuristic of the research program (Lakatos, 1978a). When such an intimate connection exists between auxiliary hypotheses and substantive theory, as is the case in physics, a challenge to the auxiliary necessitates a revision of the substantive theory. This, I believe, is what makes this kind of ad hoc theorizing acceptable for Lakatos. In psychology, unfortunately, ad hoc challenges to the auxiliary hypotheses are often little more than afterthoughts that do not have any real consequences for the substantive theory. The difference between psychology and physics in terms of the interrelations between substantive and auxiliary theories can thus account for Lakatos's negative evaluation of ad hoc theorizing in psychology.

**Null Hypothesis Tests and the Slow Progress in Soft Psychology**

In his 1978 essay, Meehl complained about the “slow progress in soft psychology.” Most theories in soft psychology, Meehl wrote, “suffer the fate that General McArthur ascribed to old generals—they never die, they just fade away” (Meehl, 1978, p. 807). Serlin and Lapsley took issue with Meehl’s pessimistic observation. They maintained that, according to Lakatos, progress takes on a historical character and thus can be detected only with hindsight. They suggested that Meehl was perhaps too expectant of rapid progress because of his faith in the Popperian reconstruction of science, which is conceived as involving a rapidly developing and continuing series of “conjectures and refutations”... where a single negative instance can purge a theory from further consideration. (Serlin & Lapsley, 1985, p. 79)

Note, however, that Meehl did not only complain that psychological theories are never conclusively rejected. Meehl’s point was that theories in soft psychology become neither accepted nor rejected; they simply become unfashionable and “fade away,” or as Meehl (1978) preferred to put it, “people just sort of lose interest in the thing” (p. 807). It is not just that theories persist in the face of anomalies or that researchers get involved in extensive revisions of their auxiliary hypotheses. For Meehl, the essence of that slow progress is the lack of accumulative knowledge and the feeling “that in soft psychology theories rise and fall, come and go, more as a function of baffled boredom than anything else” (Meehl, 1978, p. 807).

Meehl (1978) implicated the practice of null hypothesis testing as a major factor contributing to this slow progress:

I suggest to you that Sir Ronald [Fisher] has befuddled us, mesmerised us, and led us down the primrose path. I believe that the almost universal reliance on merely refuting the null hypothesis as the standard method for corroborating substantive theories in the soft areas is a terrible mistake, is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology. (p. 817)

Meehl was not very clear as to the causal relationship between null hypothesis tests and slow scientific progress. On the basis of the preceding paragraphs, however, we can now derive our own conclusions in this regard; basically, I suggest that null hypothesis testing and related practices seem to replace good theory building in psychology.

At least three mechanisms by which null hypothesis testing is liable to replace serious theory building can be pointed out. The first, and most important one, is that instead of demanding a high level of logical consistency, explanatory power, and accurate predictions from their theories, psychology researchers are trained to feel satisfied when a relevant statistic (e.g., a correlation coefficient) is statistically different from zero at the .05 level. This training is reflected in the emphasis that graduate programs in psychology put on statistics courses as opposed to courses that teach the philosophy of science or principles of scientific methodology. I can personally testify that at the urging of my professors I have learned quite a bit on sophisticated techniques to test the null hypothesis in the course of my graduate studies; there was much less of a demand that I should know the criteria for a good theory or understand the philosophical principles underlying scientific methodology.

I believe that scientific psychology may be paying a dear price for this bias in training: When passing null hypothesis tests becomes the criterion for successful predictions, as well as for journal publications, there is no pressure on the psychology researcher to build a solid, accurate theory; all he or she is required to do, it seems, is produce “statistically significant” results. It is indeed encouraging that in spite of this lack of pressure some truly powerful theories and research programs can be found in psychology, as Serlin and Lapsley (1985) and, more recently, Gholson and Barker (1985) have pointed out. I contend, however, that these are still the exception, particularly in “soft” psychology. It is no accident, I believe, that when these authors cited examples of research programs, their examples were not from social, personality, or clinical psychology.

A second mechanism that ties weak theorizing to null hypothesis testing is related to the first one. It is the illusion that these tests create in regard to the validity of the theory that is supposedly being tested. In its most extreme form, this “valid research hypothesis fantasy” becomes the “belief that statistical significance directly reflects the probability that the research hypothesis is true” (Carver, 1978, p. 386). The psychology researcher is trained to feel very comfortable about his or her theory if the experimental results were statistically significant at the .05 level and to feel quite ecstatic if they were “significant” at the .001 level. In most cases, however, this feeling is totally unjustified, as has clearly been shown by Meehl (1967) and others (e.g., Morrison & Henkel, 1970).

A third mechanism was recognized by Lakatos himself. It is the unfortunate fact that the “high-tech” statistical tests seem to create an air of scientific respectability that often passes for the real thing, at least among psy-
chologists (see Bakan, 1967). Thus, psychology researchers indulge in highly sophisticated significance tests, using computers that are accurate to the eighth decimal point, and feel that they are involved in an extremely exact science. The actual predictive power of their theories, however, can still be extremely poor.

**Do We Really Need Null Hypothesis Tests?**

Although Serlin and Lapsley emphasized that our theory testing should be made more severe, they maintained that, in principle, the practice of null hypothesis testing—that is, setting up a "straw man" competitor to the theoretical hypothesis—is required by "logical considerations." This is because the *modus tollens* can only be directed at the hypothesis that is complementary to the research hypothesis, that is, the null hypothesis. Serlin and Lapsley insisted that the physicist, in testing his or her research hypotheses, also "must set up the 'straw man' logical complement of the theoretical prediction" (Serlin & Lapsley, 1985, p. 81).

The formal rules of logical inference, however, rarely reflect the way scientists actually operate. This, in fact, is a central theme in Lakatos's philosophy, which Serlin and Lapsley claim to represent in their article. Lakatos's philosophy emerged as a response to philosophies that attempted to apply the rules of logic to scientific inference—philosophies that failed, as Serlin and Lapsley recognized, to describe the actual development of science. Lakatos's view, as well as the views of other historically minded philosophers, rejected the assumption that scientific standards could be set a priori (see Toulmin, 1976). Lakatos's philosophy was guided by the actual practice of scientists in developed sciences; it was a rational reconstruction of the history of science. And this perspective is reflected more faithfully, in my opinion, in one of Rozeboom's (1960) strongest arguments against null hypothesis tests:

> Finally, if anything can reveal the practical irrelevance of the conventional significance test, it should be its failure to see genuine application to the inferential behavior of the research scientist. Who has ever given up a hypothesis just because one experiment yielded a *p* value above .05? In fact, the reader may well feel undisturbed by the charges raised here against the traditional null hypothesis decision procedure precisely because, without perhaps realizing it, he has never taken the method seriously anyway. (p. 424)

Meehl's (1978) observation that physicists rarely perform statistical tests is also critical from the Lakatosian point of view; it strongly suggests that a good science has no need for such techniques. In fact, when the experimental results are clearly as predicted, concern with rejection of the null hypothesis seems a bit absurd: "With the corpus delicti in front of you, you do not say, 'Here is evidence against the hypothesis that no one is dead.' You say, 'Evidently someone has been murdered.'" (Berkson, 1942, p. 326). An additional observation of Meehl, that the more "noble traditions" in scientific psychology were developed "with negligible reliance on statistical significance testing" (Meehl, 1978, p. 817, emphasis in original), may also indicate that giving significance tests such a central role in psychological research, the science of psychology is stressing the marginal aspects of theory development and testing at the expense of the central ones.

How, in the absence of statistical significance tests, can a researcher establish the results of an experiment as consistent, or inconsistent, with the theoretical prediction? The controversy in the social sciences over the need for statistical tests, and for discrete decision rules in general, has been around for a long time (see Morrison & Henkel, 1970, for a collection of relevant articles). Several writers who questioned the need for statistical tests have suggested alternatives to them. Some authors recommended doing statistical tests but without granting them decision power; Skipper, Guenther, and Nass (1967), for example, recommended that the researcher should report the actual level of "significance" and "let the reader decide whether for his purpose it has any practical significance" (p. 160). Other writers (e.g., Selvin, 1957) suggested that, at least in some contexts, significance tests should be eliminated completely.

Lakatos did in fact acknowledge the problem himself and suggested a solution that has been recommended by many writers (e.g., Lykken, 1968; Stevens, 1971) but is rarely practiced in psychological research: replication (see Lakatos, 1978a, p. 107). He also reminded us of "Popper's important qualification that a basic-statement has no power to refute anything without the support of a well corroborated falsifying hypothesis" (Lakatos, 1978a, p. 24). Lakatos thus prescribed another important criterion for assigning a truth value to an observation, a criterion that is routinely used by physicists when evaluating an experimental result and its impact on the theory—that is, the strength of the theory that predicted it. Labovitz (1968) also endorsed this criterion, suggesting that in order to evaluate experimental outcome, one should examine the theoretical support of the predictions and the consistency of the findings with concurrently accepted knowledge. At any rate, arbitrary and isolated statistical decision rules are clearly not a part of the Lakatosian framework.

**Conclusion**

Several writers have offered solutions to the problem of weak theory testing in psychology (e.g., Nunnally, 1960; Swoyer & Monson, 1975). The basic idea in most of these is increased emphasis on effect size in prediction and testing. Serlin and Lapsley's (1985) article offered the most complete and sophisticated of these solutions to date. I have argued here, however, that Serlin and Lapsley's attempt, at the same time, to defend the rationality of the actual practice of psychologists was unwarranted. By appealing to Lakatos's methodological falsificationism, Serlin and Lapsley concluded that soft psychology is in fact progressive and that, in principle, null hypothesis testing...
and other practices criticized by Meehl constitute rational scientific strategy. As shown here, however, this conclusion is a bit too generous and involves too liberal an interpretation of Lakatos's philosophy. Lakatos himself was careful to emphasize that although his philosophy was indeed more liberal than Popper’s, it was also more strict in that it demands not only that a research programme should successfully predict novel facts, but also that the protective belt of auxiliary hypotheses should be largely built according to a preconceived unifying idea, laid down in advance in the positive heuristic of the research programme. (Lakatos, 1978b, p. 149)

And lest the reader question the relevance of these words to psychology, Lakatos added in a footnote, “I called patched-up developments which did not meet such criteria ad hoc stratagens…. a particularly good example is Meehl’s anomaly, i.e., Meehl’s (1967) critique” (Lakatos, 1978b, p. 149).

I suggest, then, that a more balanced interpretation of Lakatos’s philosophy would lead to the rejection, in principle, of null hypothesis tests or any other form of arbitrary, automatic decision rule. It would also lead to the conclusion that only after one has created a bold theory that can be meaningfully tested can one legitimately “count noses” and use ad hoc explanations, including appeals to auxiliary hypotheses, to account for experimental failures.

Finally, I would like to point out that the validity of the arguments presented in the present article does not hinge on accepting Lakatos’s point of view in regard to scientific psychology. Lakatos’s philosophy may certainly have its shortcomings, as was argued in a recent review (Gholson & Barker, 1985); the point of the present essay, however, is that Lakatos’s philosophy cannot be interpreted as defending the rationality of psychological research methodology, and particularly of null hypothesis tests. I am concerned that Serlin and Lapsley’s (1985) article will be interpreted as a philosophical legitimization of our ongoing scientific practices, coupled with too minor a call for stiffening the experimental hurdle. My belief is that the message should be different: that we should direct our energy into building integrated research programs capable of producing strong and meaningful predictions and contributing to real scientific progress in psychology. In our current state of development, I believe that statistical significance tests and related practices are often serving as an unfortunate substitute for a truly creative and progressive scientific enterprise.

REFERENCES


