The Journal of Mind and Behavior
Autumn 2001, Volume 22, Number 4
Pages 389–406
ISSN 0271–0137

The Affiliation of Methodology with Ontology in a Scientific Psychology

Matthew P. Spackman and Richard N. Williams

Brigham Young University

The misconception that the application of statistical methods makes psychology a science is examined. Criticisms of statistical methods involving issues related to the generalization of aggregate-level findings to individuals, the impoverished language of numbers, the application of questions to methods, and the logic of statistical hypothesis testing are reviewed. It is not suggested, however, that statistical methods be abandoned. Instead, it is suggested that shortcomings of statistical methods indicate the importance of making ontological considerations a primary concern. Methodological considerations in the absence of an understanding of the truth or ontological status of what is being studied will inevitably undermine psychologists' efforts at understanding what it is to be human. Whereas the use of statistical methods in psychological research does not make the discipline a science, the *truthful* affiliation of methodology with ontology may.

Nothing in the historiography of psychology is more fundamental to our disciplinary self-image than our achieving the status of science (see Danziger, 1990; Gigerenzer, 1987; Gigerenzer, Switjink, Porter, Beatty, and Kruger, 1989; Koch, 1961; Robinson, 1995). This status, so the accounts go, was won on the basis of hard-fought, conceptual and methodological battles. The end result of this process was that we not only severed the ties that bound us to philosophy, but produced empirical results that demonstrated the advantage of experimentation over speculation and *mere* rational argument.

It is not our purpose to recapitulate this history. We propose instead to examine a related topic upon which that history depends, a fundamental misconception we believe to be held by many psychologists: that application

Requests for reprints should be sent to Matthew Spackman, Ph.D., Department of Psychology, Brigham Young University, 1001 Spencer W. Kimball Tower, Provo, Utah 84602. Email: matt_spackman@byu.edu

of a particular methodology — statistical methods[1] — makes "scientific" the various research programs within psychology. This is simply not the case. Methods do not make disciplines scientific. Rather, we suggest that the status of "science" may be appropriately claimed by disciplines which have revealed significant truths about their subjects of study. It is not, then, methodological advance which makes a discipline scientific, but ontological revelation.

We first review four criticisms of statistical methods which have, for some time, received attention in the literature. These criticisms involve issues related to the generalization of aggregate-level findings to individuals, the impoverished language of numbers, the application of questions to methods, and the logic of statistical hypothesis testing. By reviewing these potential pitfalls of mainstream psychology's adopted method, we hope to remind readers of the naiveté of rote application of statistical methods to psychological questions and thereby establish the importance of making ontological considerations a primary concern to any research program. We do not, however, advocate the abandonment of statistical methods in psychological research. Instead, we advocate an informed and self-critical methodology, one dependent upon, or affiliated with, primarily ontological concerns.

Means Don't Tell Us About AnyONE

Early in the history of the use of probability theory in the social sciences, a controversy arose as to what exactly the probability of a group's behaving in a certain fashion might imply for the behavior of particular individuals in the group of interest (Danziger, 1990; Gigerenzer et al., 1989; Lamiell, 1998; Porter, 1986). For example, if it is the case that, in any given year, nationwide, 80 percent of psychology professors receive a five-percent raise, what would the chances be for either of the authors of the present paper to receive such a raise this year? Nineteenth-century probabilists debated whether such demographic data were evidence of natural laws. Two alternative interpretations of such laws were offered. According to the first interpretation, the laws evidenced by the data held at the aggregate *and* individual levels. In other words, 80 percent of all psychology professors might expect a five percent raise this year, and the present authors would each have an 80 percent chance of receiving a raise this year. According to the second interpretation, the laws supposedly evidenced by such demographic data held at the aggre-

[1]By the term "statistical methods" we mean not simply the formulas and procedures researchers apply to their numeric data, but also the means or methods by which they obtain those data, especially when those methods are designed to yield data suitable for analysis in a particular statistical test. Another term we could have chosen is "empirical methods." This term is not, however, so indicative of the bias in methodological choice we address with the term "statistical methods."

gate, but not at the individual level. In other words, 80 percent of all psychology professors might expect a five percent raise this year, but, much to the chagrin of the present authors, their chances of receiving a raise this year would depend upon their individual circumstances.

Pertaining to the practice of research in psychology, the above debate over the level at which laws may be applied becomes an issue of methodological and theoretical importance. Danziger (1990) outlined three methodological traditions in the development of psychology as a modern science: the Leipzig, clinical, and Galtonian models. The Leipzig model was pioneered by Fechner, Wundt, and those that followed them in the investigation of psychological principles at the individual or idiographic level. This model sought to find regularity at the individual level in an attempt to outline psychological laws that held at the aggregate level (Lamiell, 1995, 1998). In this way, the Leipzig, or idiographic model resembles the first of the interpretations of the authors' salary expectations explained above — the idea that the identified laws hold at both aggregate and individual levels.

The clinical and Galtonian models were similar to one another in their differences from the Leipzig model. Both the clinical and Galtonian models favored the analysis of their subject matters at the aggregate level. Those adhering to the clinical model studied groups of persons that might be categorized under a specific diagnosis. Those that adhered to the Galtonian model were interested in how variables such as IQ and task performance might co-relate, an analysis performed by examining aggregate-level data. Both of these models could be considered nomothetic and are related to the second interpretation of natural laws described above, the idea that the identified laws hold at the aggregate, but not the individual level.

The essential difference between the idiographic and nomothetic research traditions is that the idiographic tradition looks for consistency at the individual level as evidence for aggregate laws, whereas the nomothetic approach looks for consistency at the group level as evidence for aggregate laws (Lamiell, 1998). It is the nomothetic method's ignoring of variation at the individual level that makes this method particularly inapt for a scientific discipline. This is because when a regularity is found at an aggregate level, what is actually found is a probabilistic law that says nothing about the individuals in the aggregate to which the law is said to apply (Danziger, 1990; Lamiell, 1987, 1995, 1997, 1998; Lamiell and Durbeck, 1987; Valsiner, 1986).

The mathematical basis for such a critique is simple. All inferential statistics are based on the analysis of means, variances, or proportions. When a test statistic is calculated, it is derived solely on the basis of how an aggregate statistic compares to the same statistic derived from another group. Each individual's score is important to the calculation of the test statistic, but conclusions are never drawn from an individual's data. They are based on derived values and are not (necessarily) the actual scores of any individual in the sample. While such inferential procedures are an attempt to identify laws that affect everyone (nomothetic laws), they actually inform us about no one.

Proponents of a nomothetic approach to psychological research (an approach which relies upon aggregate-level regularities or probabilistic laws) might suggest that such condtional laws are all that can be hoped for at present, given the complexities of human behavior. In other words, such proponents would suggest that (and this is a position which dates back to at least J.S. Mill's "empirical laws," or laws of regularities, see Robinson, 1995), whereas causes of individuals' behavior cannot be identified, we can identify probabilities for individuals to engage in behaviors from aggregate-level data.

The shortcoming of any probabilistic statement, and the reason that the findings of such statistical regularities cannot qualify a discipline as scientific, is that they are descriptive observations masquerading as causal explanations. The essential logical proof for how the total proportion of professors receiving salary increases can cause or explain any particular professor's salary increase is missing. Conditional probabilities of the sort *p(raise | number of publications, committee service hours, etc.)*, as well as Bayesian subjective probabilities, fare no better. Both add detail to the calculation of probabilities, but come no closer to identifying a cause or explanation of an event.

The Impoverished Language of Numbers

Imagine a researcher studying love in long term romantic relationships. Conventional wisdom might call for measuring the strength of subjects' love for their partners with one or more nine-point, Likert-type scales, where 0 might be labeled as "not much" and 8 as "a lot." We would ask, however, whether a rating of 8 on such a scale (or the result of a linear combination of multiple scales) adequately conveys to the researcher how much a particular subject loves his or her partner. This is a problem of validity. However, the issue is a bit more complex than is usually acknowledged.

We suggest that such a rating cannot adequately describe subjects' love for their partners (Danziger, 1990; Kvale, 1994; Williams, 1999). While numerical data may have certain properties perceived as being advantageous to psychological researchers, the disadvantages far outweigh any perceived advantages. Steinar Kvale (1994) lists four reasons why researchers might choose to translate human experience into numeric form. First, it may be the case that researchers hold an ontological assumption that the world itself is actually mathematically ordered, or that it offers itself to persons in an inherently numeric fashion. Surely, certain aspects of the world such as age, the final score of a game, and distances do offer themselves in such a way. Many aspects of our lives, however, and quite possibly the most meaningful aspects, are not so easily quantified. Second, there may be an epistemological demand for research findings to be commensurate. We might imagine the difficulties the researcher in the example above would have in summarizing and presenting his data in any conclusive manner if he, rather than employing Likert-type scales, had asked respondents to write one-page essays on their feelings for their partners. If one respondent indicates that he "really" loves his partner, while another indicates that she loves her partner "very much," which one loves his or her partner more? This problem of commensurability is seemingly avoided with a numeric scale. All one has to do is assume that an 8 for one respondent means exactly what it does for another respondent. Third, it may be that there is a technical interest in numerizing one's data. Our researcher could run a simple two-by-two ANOVA on the love ratings and create a graph of the interaction which would summarize all of his findings. If the researcher had asked for the essays described above, however, he would have difficulty in trying to summarize numerous one-page essays in a short journal article addressing all of the relevant theoretical issues. This difficulty is a staple in discussion of the relative merits of quantitative vs. qualitative research strategies. In the end, however, it may come down to an issue of rhetoric. Few in the discipline, it seems, would dispute the superior rhetorical power of a factor analytic technique over a merely verbal summary of words such as "really" and "very much." To many psychological scientists, it appears that numbers are more convincing than words.

Some of the limitations of numbers were hinted at in the above elaboration of the reasons for which researchers may choose to employ numeric data. Numbers can convey only three sorts of information: frequency or numeric count, numeric magnitude, and numeric pattern or co-relations among sets of numbers. As a language, numbers can capture and convey certain types of information efficiently and accurately. A numeric language is, however, inapt for expressing any other type of information. This sentiment is echoed in Helen Sullivan's (1952) cautionary statement on employing the mathematical method:

> When the mathematical method, which is valid in itself, pronounces on the ultimate nature of things, it, *by that very fact*, poses as a philosophy; it usurps the role of metaphysics. It cannot, therefore, ignore two of the four fundamental ontological causes [formal and final (see Robinson, 1995)] and feel that it has done its duty. (p. 68, emphasis in the original)

Lacking connotative meaning, context, or nuance of any sort, numbers are ideal for analytical purposes; these very qualities, however, make them difficult to apply to the human sphere (Slife and Williams, 1995). In fact, as soon as numerical measures are applied to inherently contextual and meaning-making human beings, those very abstract and numerical properties that make numbers so useful in analysis disappear. In their role as representations of the world of human beings, numbers are distinctly impoverished compared to more natural languages human beings create and appropriate. The precision, lack of ambiguity, and clear definitions that characterize numbers *qua* numbers do not apply to numbers as representations of the meaningful world of human beings.

In asserting the superiority of words over numbers in describing the human condition, we are not suggesting that verbal descriptions avoid all difficulties in conveying information. Our everyday experiences in searching for the right terms in which to adequately express ourselves witnesses to the inherent difficulties of expressing ourselves in any language. We do suggest, however, that verbal descriptions of at least many of the most important aspects of our lives have a distinct advantage over numerizations of those descriptions. They do not suffer from inherent difficulties in translation. (This issue of difficulties in translation continues to be important in our discussion of hypothesis testing below.)

Apart from basic measures such as rates and frequencies, numerical psychological data have as their original form, or language, words. There is, then, in most cases, a necessity for researchers (e.g., in the case of coding) or subjects (e.g., in the case of traditional Likert-type scales) to translate words into numbers. The degree to which an idea may be translated from one language to another is dependent upon two things: the skill of the translator and the limitations of the language into which the original idea is translated. While proponents of numerization might suggest that a skillful translator (i.e., someone skilled in measurement) can overcome the limitations of the language of numbers, we disagree. Whereas a numerical language is particularly apt for measurement (i.e., measurement of equivalence, quantity, and pattern), much of human experience is not quantity-relevant. A science of the human condition must be able to explain human behavior. It must be able to transact in reasons explanations in the tradition of William Dray's (1957) historical explanation. Such explanations do not rest on the measurement of quantities, but upon the quality of contextual circumstances. Numbers alone cannot, therefore, serve as the language of these accounts. Only words (but here, too, with difficulty) can begin to serve researchers' needs in explaining human actions.

By this point it should be apparent that many critics of the numerizing of psychological data recommend psychologists employ qualitative, rather than quantitative, analytic techniques. Whereas the collection and analysis of qualitative data is becoming increasingly popular in other social sciences, psychologists have not been eager consumers or practitioners of qualitative research methods (Denzin and Lincoln, 1994). We grant that much of the hesitancy psychologists experience with regard to qualitative methods may be based on documented shortcomings of self-report data (see Nisbett and Schachter, 1966; Nisbett and Wilson, 1977; Wilson, 1985). It has been shown, for example, that persons are often inaccurate in their explanations of their actions (Wilson, Laser, and Stone, 1982), predictions of future actions (Shrauger, Ram, Greninger, and Mariano, 1996) and feelings (Loewenstein and Schkade, 1998), as well as in other areas of self-perception (Wilson, Dunn, Kraft, and Lisle, 1989). It should be noted, however, that the difficulties of self-report data are not limited to qualitative data, but are also manifest in numeric self-reports. Certainly measurement theorists have developed methods for addressing many of these concerns (see Nunnally and Bernstein, 1994), but the fact remains that we are often simply inaccurate in our reports of our activities — regardless of the language in which we are asked to respond. And so, though we advocate the use of qualitative methods in psychological research, we do not reject the use of quantitative methods. The language of numbers and the language of words each have their own strengths and limitations. We suggest researchers use the language called for by their understanding of the ontological status of their subjects of study, as will be described below.

Applying Methods to Questions

It has been suggested that psychologists adopted statistical methods in an attempt to achieve scientific credibility (Danziger, 1990; Gigerenzer, 1987; Gigerenzer et al., 1989; Koch, 1961; Robinson, 1995, 1998, in press). The application of mathematical principles to scientific endeavor has traditionally been equated with progress in the discipline (Robinson, 1995, 1998, in press). In the words of Richard Schlagel (1995), "Except for the description and the classification of phenomena, and the verbal formulation of inductive generalizations or laws, advances in the physical sciences have gone hand in hand with developments in mathematics" (p. 125). It is worthy of note, however, that the mathematizing of knowledge is not a matter of applying statistical methods. Statistical methods seek to model, not to explain. Statistical methods seek to quantify doubt — they do not solidify certainty. While statistical methods have advanced the study of psychological phenomena, that progress has not come without the adoption of a particular worldview.

The most important ramification of this limited perspective underlying the application of statistics to psychological research is that it limits researchers in the types of questions they may address in their research, as well as the methods they may employ to study phenomena. (We offer an example of how methods may constrain research in our comparison of two approaches to anger, below.) Gigerenzer et al. (1989) point out that one of the great strengths of statistical methods is that such "techniques . . . do not inhere in any concrete subject matter — and for that reason are breathtakingly general in application" (p. 274). While this statement is true, certain statistical tests are more widely used and understood than others, and certain data will lend themselves more readily to such tests. For example, the two-by-two ANOVA might be considered the workhorse of social psychology. Its two possible main effects are easily understood, and its possible interaction may be simply plotted. Given the intuitive appeal of this design, it is easy to see how a researcher wishing to analyze her data in this way should design her study. All that is needed is to extract a single numeric measure from subjects in one of four treatment combinations.

There is something potentially misleading about this process of designing a study before it is clear and settled just what is being studied. The problem lies in matching questions to methods rather than methods to questions.[2] What is easily overlooked is the fact that many psychological phenomena may not naturally manifest in any set of four conditions created by statistical design. They may not be captured by a single measure. They may also not be representable in a numeric framework at all.

That the problem we are describing here is a genuine one is illustrated in the following description. Anecdotal evidence suggests to us that it is not uncommon for researchers to approach a statistical consultant for assistance in analyzing their data. Not infrequently the consultant expresses the opinion that the experimental design could have been "improved" so that a particular statistical analysis might have been applied. From the perspective endorsed here, however, we might ask why a researcher should formulate her measuring instrument around a statistical analysis. It seems that psychologists have, in some cases, yielded control over the conceptually crucial design and measurement aspects of their research programs in subservience to purely statistical considerations (Danziger, 1990; Gigerenzer et al., 1989; Robinson, 1995, 1998, in press; Slife and Williams, 1995; Williams, 1999). Rychlak (e.g., 1981, 1988) has argued persuasively against such a confounding of theories and methods. His articulation of the problem of the "S–R bind," in which concern for an allegiance to particular methods dictates which types of theories can be formulated and tested, serves as an example of our caution and as a warning to the discipline.

One suggestion offered for resolution of the dilemma of inappropriate assignment of statistical methods to this methodological role is that researchers should first outline general hypotheses they are interested in studying, and then match appropriate research methods to them (Robinson,

---

[2]As will be described, the matching of methods to questions is preferable to a naïve methodological eclecticism, but still deficient in certain respects compensated for by what we describe as an affiliation of methodology with ontology.

1995, 1998, in press; Slife and Williams, 1995; Williams, 1999). Doing otherwise runs the risk that methods themselves will determine what we study, as well as our results. Contrary to the sentiment expressed by Schlagel above, Robinson (in press) suggests that:

> Progress in science arises from the application of an informed imagination to a problem of genuine consequence; not by the monotonous application of some formulaic mode of inquiry to a set of quasi-problems chosen chiefly because of their compatibility with the adopted method.

Our own stance on the relation of methods to questions is sympathetic to the recommendations of Robinson and others. We suggest, however, that prior consideration of the ontological status of the subject of study is essential to any attempt at matching methods to questions.

The Logic of the Null Hypothesis

While not all statistical techniques involve the testing of hypotheses, the majority of them do, and the validity of such methods for the evaluation of theories has been criticized (Bakan, 1966; Chow, 1996; Cohen, 1994; Dar, 1987; Hagan, 1997; Harré, 1972; Meehl, 1978, 1990a, 1990b; Serlin and Lapsley, 1985). These criticisms revolve around the issue of what conclusions may be drawn on the basis of testing the null hypothesis.

When the null hypothesis is not rejected. It is typically assumed that, when the null hypothesis is not rejected, the treatment has had no effect, or that the theory being tested has been falsified. However, this would be an assumption without logical necessity (as most of us learned, and then quickly forgot, in our introductory methods courses), and the question as to what exactly to conclude, an issue not yet resolved by philosophers of science (Meehl, 1978, 1990a, 1990b). The logical form for the test of a given theory T is as follows [see Meehl, 1990a]: $(T \cdot A_i \cdot C_p \cdot C_n) \rightarrow (O_1 \supset O_2)$, where T is the theory to be tested, A_i represents the collection of auxiliary theories necessary to the testing of T, C_p is the *ceteris paribus* or "everything else being equal" clause, C_n is the experimental conditions, and $(O_1 \supset O_2)$ signifies the conjunction of experimental observations. If our results falsify the right hand conditional portion of the logical statement above, then it is held that the left hand conjunction is falsified as well, amounting to Popper's (1959) *modus tollens* (though, see Cohen, 1994; Hagan, 1997, for discussion of whether null hypothesis tests actually constitute *modus tollens*).

However, such a falsificationist approach neglects the fact that falsifying the right hand conditional necessarily falsifies the *conjunction* of the left hand statement, which does not amount to a direct test of the theory. The researcher does not know whether the theory is false, one or more of the auxiliary theories is false, there is a problem with the experimental conditions, or whether the problem lies with the *ceteris paribus* clause. When a null hypothesis is not rejected, the question then becomes what part of the conjunction is suspect.

Imre Lakatos (Lakatos and Musgrave, 1970) demonstrated that, in the practice of science as reflected in its developmental history, failure to obtain support for theories rarely results in researchers abandoning their theories and declaring them falsified (Dar, 1987; Meehl, 1978, 1990a, 1990b; Serlin and Lapsley, 1985; Worrall and Currie, 1978). Instead, some other element of the conjunction is declared at fault and explanations for non-rejection are generated. It is apparent that findings can falsify a theory only if they constitute the crucial test of the theory. The falsificationist approach can never constitute such a test.

It is this deficiency in null hypothesis testing that compromises the researcher's ability to engage the scientific process. Null hypothesis testing does not constitute a means for theory testing because it is not the theory itself that is being tested. The procedure does not outline a process by which researchers can know whether a theory has been sufficiently shown to be false. In Lakatos' terms, when a research program has required too much in the way of ad hoc explanation to justify its continued pursuit, it has reached a point of "degeneration," and ought to be abandoned (Dar, 1987; Meehl, 1978, 1990a, 1990b; Serlin and Lapsley, 1985; Worrall and Currie, 1978). However, it has proven impossible to establish means for determining whether a theory is degenerate. It is this impossibility of showing theories to be false by way of null hypothesis testing procedures that makes the sole use of traditional statistical methods inappropriate to a scientific psychology.

When the null hypothesis is rejected. Not only are null hypothesis testing procedures inadequate to the task of theory-testing when the null hypothesis is not rejected, but there are also inadequacies when the null hypothesis is rejected. The first deficiency of the procedure is a logical fallacy committed by researchers employing null hypothesis procedures. The above logical form of the test of theory T indicates that when results support the conditional ($O_1 \supset O_2$), the researcher concludes the truthfulness of ($T \,.\, A_i \,.\, C_p \,.\, C_n$), or, more usually, of simply T. When applied to tests of the null hypothesis, this amounts to the formal fallacy of denying the antecedent, which takes the form of if p then q; not-p / ∴ not-q (see Audi, 1995; Slife and Williams, 1995).

Related to denying the antecedent is the question of just what it is that is known or supposed to be true when the null hypothesis is rejected. Failing to reject the null hypothesis leads to the dilemma of what aspect of the conjunction of to-be-tested theory, auxiliary theories, *ceteris paribus*, and experimental conditions has been falsified. This same dilemma arises when the null hypothesis is rejected. We must inquire as to what is producing the experimental difference found by our statistical procedure. Meehl (1990a) suggests that it is likely that the theory itself is corroborated because the other elements of the conjunction are derived from it. This may be the case, but is not a logical necessity. As the debates surrounding the Duhem–Quine thesis have demonstrated, theory is always underdetermined by data (see Curd and Cover, 1998, for a review).

Another shortcoming of the null hypothesis testing procedure is that, as Chow (1996) has noted, the statistical hypothesis being tested and the theory presumably being tested are not identical. Given that this is the case, a researcher makes a conceptual leap in interpreting her rejection of the null hypothesis as evidence for the truth or adequacy of her theory. The process of formulating the null hypothesis involves the translation of the to-be-tested theory at a number of levels. In order for the rejection of the null hypothesis to constitute corroborative evidence for the theory, the translation at each level must be shown to be a 1:1 translation. In other words, a statistical hypothesis must be exactly equivalent to an experimental hypothesis, which must be equivalent to a research hypothesis, which must be equivalent to a substantive hypothesis, which must be equivalent to the theory in question. In the absence of such equivalence, we have conjecture, but not evidence. In essence, unless an experimenter's theory is directly operationalizable as a statistical hypothesis, one never has direct access to a theory through one's testing procedure (Meehl, 1978, 1990a). It should be noted, however, that any theory that is directly operationalizable is not really a theory at all because it does not extend beyond the empirical conditions of which it is entirely composed. Furthermore, any test of statistical significance merely affirms at the empirical level that a pattern of numerical results is improbable given certain purely mathematical assumptions. It can never provide a probability estimate of the truth or falsity of an explanation for that pattern of numbers.

It must be granted that the problems of affirming the consequent, the Duhem–Quine thesis, and operationalization are inherent to all (or nearly all) psychological research that seeks to validate theories, including those employing qualitative data. Given that this is the case, what are we to do? In support of Meehl's (1967, 1978, 1990a, 1990b) and others' (Hagan, 1997; Lykken, 1968, 1991; Serlin and Lapsley, 1985) recommendations, we offer the following. Although experimental data can never constitute direct falsification or corroboration of theory, they can afford justification for placing confidence in a theory. If enough data thus accrue, we may choose to accede and accept the theory as worthy of continued confidence and further work. In the absence of sufficient support, we will withhold our ascent. What constitutes strong or convincing support for a theory? As Meehl suggests, support for a particular theory will be convincing to the degree that predictions generated by the theory would be surprising if the theory were false. As many such coincidences accumulate, support for the theory may become so great as to prompt us to conditionally accept it.

We want to stress here that support for theories accrues as a series of successful demonstrations and is most convincing when that support is generated by a variety of methods (Meehl, 1978, 1990b; Slife and Gantt, 1999). Here we are in agreement with many qualitative researchers who advocate the idea of "triangulation," or the evaluation of a theory in terms of data gathered from multiple sources, varying perspectives, and diverse methods (c.f., Denzin and Lincoln, 1994). In addition to multiple designs employing traditional null hypothesis testing procedures, we advocate also the use of other quantitative methods such as multivariate techniques which do not rely on traditional null hypothesis testing procedures (e.g., cluster analysis, factor analysis, discriminant analysis, etc.), as well as qualitative methods. We advocate a genuine methodological pluralism where statistical hypothesis testing does not supercede other procedures or dictate the course of the scientific process. Our advocacy of methodological pluralism is, however, conditional in the sense that we support pluralism only when the methodological considerations inherent to such a process are affiliated with, or are in the service of, primarily ontological considerations, as described below.[3]

The Relationship of Method to Science

From the foregoing, it should be apparent that the use of statistical methods in psychological research is problematic in a number of respects. It should also be apparent, especially in our recommendations regarding the evaluation of theories, that we do not suggest that statistical methods be abandoned. We submit what may seem like a radical proposition, that the use of statistical methods in psychological research does not make our discipline a science, but that the *truthful* affiliation of methods with ontological considerations may.

Scientific enterprises employ methods, and such methods compared across sciences may indeed have commonalities. However, a science is not a science merely because it embraces a particular method and deploys it in the study of phenomena within its domain. What makes an endeavor a science is a particular sort of knowledge claim. Psychology will have a supportable claim to the status of science when it offers a type of knowledge about the human condition which is not merely possible and coherent, but compelling and enlightening.

Hans-Georg Gadamer (1982) suggests that, contrary to the faith of our modern world, and what appears to be the prevailing presumption of social sci-

[3]By "ontology" we mean the researcher's view of the reality of what she is studying, her understanding of the to-be-studied phenomenon and its existence in the world.

ence, truth (or, scientific knowledge) does not necessarily result from the application of method. Rather, it must be the case that all methods, in their development, deployment, and products depend on prior understandings of truth, that is, on ontological assumptions. In simplest terms, methods are developed in such a way that they appear to be effective at getting at truth — but at truth as we suppose it to be. We who craft the method — including statistical methods — must surely have some conception of truth in order to know what the method should do, what it should ignore, and to what it should be sensitive. Developing a method is not unlike constructing a measuring instrument, such as a barometer. In order to build a barometric, we must first know, or at least assume, something about atmospheric pressure so that we can build a barometric sensitive to atmospheric pressure as we understand it.

Methods, like instruments, can only reflect back to us the information to which they are sensitive. Thus, they verify through the results they produce whether our initial understanding of *truth* is reasonable in the face of observation, but they cannot directly reveal to us the truth of our understanding. It is for this reason that methodology must always be affiliated with ontology. Only through coherent discourse about what *is* and the nature of its *being* (in Heidegger's, 1953/1996, sense of the term) can we determine how a phenomenon may best be studied. Method should, therefore, be in the service of science, a part of its development, a necessary constituent of, but never a sufficient condition for science.

We are not alone in calling for a reformation in the relationship of methods and science, as is demonstrated in our review of the critiques of statistical methods above. Such calls for reformation have typically come at two levels. At the most basic level, what we consider to be a naïve methodological pluralism has been suggested (see McLeod, 2000; Schmitt, Couden, and Baker, 2001; Thoresen, 1999). Such a methodological pluralism would hold that no method is inherently superior to any other. In particular, adherents of methodological pluralism would claim that qualitative methods are not inferior to quantitative methods. Though such a methodological pluralism constitutes what we consider to be an advance beyond the current hegemony held by statistical methods in contemporary psychology, it is inherently deficient. Given that methods reflect prior commitments to the ontological status of the phenomena they are intended to engage, to hold that no particular method is inherently more suited to a particular phenomenon, as a naïve methodological pluralism suggests, implies that the question of the ontological status of the subject of study is unimportant.

The second call for reformation suggests that researchers match their methods to their questions, as discussed above (see Slife and Gantt, 1999; Wertz, 1999; Yanchar, 1997). This second call entails the assumption that, whereas qualitative and quantitative methods are both legitimate forms of scientific study, particular methods are more or less appropriate to particular questions, an advantage over naïve methodological pluralism's relativistic attitude toward methodology. However, this second call for reformation still does not recognize the importance of ontological issues to science. There seems to be an assumption that the questions to which methods are to be matched are the "right" questions. In other words, the necessary ontological discourse defining the nature of the object of study is, in some sense, bypassed.

Whereas we agree that quantitative methods are not inherently superior to nor more scientific than qualitative methods, and that our methods must be adapted to our questions of interest, we suggest an affiliative role of methods not entailed in either of these calls for reformation. What we are suggesting here is that, in order for psychology to obtain the status of science, our methodology must be derived from, or affiliated with, ontological consideration. Ours is a call for a methodological pluralism in which methods are developed, selected, and deployed based on the likelihood that the results of their use will be appropriate, enlightening, and useful given the prior ontological conclusions of the researcher.

We offer as an example of the importance of consideration of ontological commitments to selecting a method of study a comparison of two approaches to the study of anger, Leonard Berkowitz's and James Averill's. In his summary of his cognitive–neoassociationist theory of anger, Berkowitz (1993) defined anger as an "associative network in which specific types of feelings, physiological reactions, motor responses, and thoughts and memories" (p. 9) are activated automatically by unpleasant experiences. Berkowitz is careful to point out that, because his theory claims automaticity of the connection between noxious stimuli and the "anger syndrome," it can explain anger at such non-arbitrary and socially-legitimate occurrences as interruptions (Parrott, 1993). Berkowitz suggests that more cognitively-oriented theories would have difficulties explaining reactive angers because they suggest anger arises from a recognition of a purposeful wrong committed on oneself or a close other. Berkowitz criticizes such cognitively-oriented theories for relying on self-report data, which he suggests offers insight into how persons view or interpret their anger experiences and not into anger itself.

In this respect, it may be seen how Berkowitz's ontological commitments with regard to anger affect his methodological considerations. For Berkowitz, anger is a syndrome which is activated automatically. Because of its automaticity, it cannot be accurately assessed through self-report methods, as persons may not be aware of the process of an anger experience because its inception is below the level of consciousness. Given his ontological commitments, it is not surprising that Berkowitz has employed traditional experimental methods involving the collection of numeric data. In his well-known studies of anger (see Berkowitz, 1982, 1989, for reviews), Berkowitz has typically exposed his subjects to unpleasant stimuli (e.g., cold water, arms raised for extended periods) and then measured subsequent acts of aggression. He has, with fairly high consistency, found that such unpleasant stimuli have been associated with increases in aggressive behavior.

However, researchers committed to more cognitively- or socially-oriented approaches to anger suggest that Berkowitz's conception of anger, while largely supported by his experimental program, ignores important aspects of what it means to be angry (see Averill, 1983b; Parrott, 1993). In his own research, Averill (1983a, 1983b) has focused on how anger functions within relationships. He suggests that anger is a role persons take on when they seek to adjust their positions in social relationships. As such, Averill holds that anger does not exist as a physiologically-oriented syndrome, as Berkowitz suggests, but that it is socially constructed. Averill's ontological commitment with regard to anger may be seen to differ greatly from Berkowitz's and, not surprisingly, has resulted in a different methodological approach.

Because of his commitment to a social constructionist perspective, Averill's primary interest is in the social-situatedness of anger experiences. This leads him to a self-report methodology in which subjects describe aspects of their anger experiences. Averill found that anger did, in fact, most often result when his subjects felt they were, in some sense, wronged by someone with whom they shared a relationship, and that the anger itself served as an entrée to negotiate the perceived wrong in the relationship. Averill and others suggest Berkowitz's theory cannot account for social functions of anger. Berkowitz suggests, however, that Averill's conception of anger cannot account for just the sort of anger experiences which his own theory addresses so well: those reactive angers which seem to result from situations not necessarily social in nature.

With regard to the role of ontology and methodology in establishing psychology as a science, it is worth repeating that Berkowitz and Averill's disagreement is largely one of ontology, and not of methodology. Both researchers selected methods appropriate to their prior ontological commitments. Both research programs obtained support for their respective theories in what we would suggest to be equally scientifically rigorous studies. However, Berkowitz's and Averill's theories are, at the level of the ontological understandings of their phenomena of study, incommensurable. What is important here is that disagreement over the relative merits of the two theories of anger is at the level of ontology, and not methodology (though, in cases where a methodology is deficient in some sense, the debate may take place at this level). Debate over the relative scientific merits of cognitive–neoassociationist and social constructionist approaches to anger will not, then, be settled at the level of method, but at the level of the truth of what anger itself is. Certainly this discussion will involve conversation on the appropriateness of particular methods, but the appropriateness of any method can only be established given prior ontological considerations.

This process of questioning our choice of methods *is methodology*. Methodology thus conceived should be an integral part of the discipline of psychology. But it must be a methodology based on an underlying willingness to take on the question of the truth or ontology of our objects of study, and one open to the possibility that our disciplinary preconceptions of that truth may be in error and in need of revision. Thus, the methodological question is addressed from two directions. First, we must undertake serious examination of the nature of the world and our understanding of the truth of it. Second, we ask ourselves whether our a priori commitments to received methods may blind us to understandings that otherwise are available. This sort of methodological concern, in either of its manifestations, has not been among the salient and defining preoccupations of psychology. Our contention is that it should be.

References

Audi, R. (Ed.). (1995). *The Cambridge dictionary of philosophy*. Cambridge: Cambridge University Press.

Averill, J.A. (1983a). *Anger and aggression: An essay on emotion*. New York: Springer.

Averill, J.A. (1983b). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, *38*, 1145–1160.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*, 423–437.

Berkowitz, L. (1982). Aversive conditions as stimuli to aggression. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Volume 15, pp. 249–288). Orlando, Florida: Academic Press.

Berkowitz, L. (1989). Frustration–aggression hypothesis: Examination and reformulation. *Psychological Bulletin*, *106*, 59–73.

Berkowitz, L. (1993). Towards a general theory of anger and emotional aggression: Implications of the cognitive–neoassociationistic perspective for the analysis of anger and other emotions. In J.R. Wyer and T. Srull (Eds.), *Perspectives on anger and emotion* (pp. 1–46). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Chow, S.L. (1996). *Statistical significance: Rationale, validity, and utility*. Thousand Oaks, California: Sage.

Cohen, J. (1994). The earth is round (p< .05). *American Psychologist*, *49*, 997–1003.

Curd, M., and Cover, J.A. (Eds.). (1998). *Philosophy of science: The central issues*. New York: W.W. Norton and Company.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge: Cambridge University Press.

Dar, R. (1987). Another look at Meehl, Lakatos, and scientific practices of psychologists. *American Psychologist*, *42*, 145–151.

Denzin, N.K., and Lincoln, Y.S. (Eds.). (1994). *Handbook of qualitative research*. Thousand Oaks, California: Sage.

Dray, W. (1957). *Laws and explanation in history*. London: Oxford University Press.

Gadamer, H.-G. (1982). *Truth and method*. New York: Crossroads.

Gigerenzer, G. (1987). Probabilistic thinking and the fight against subjectivity. In L. Kruger, G. Gigerenzer, and M.S. Morgan (Eds.), *The probabilistic revolution, Volume 2: Ideas in the sciences* (pp. 11–33). Cambridge, Massachusetts: MIT Press.

Gigerenzer, G., Switjink, Z., Porter, T., Beatty, J., and Kruger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge, England: Cambridge University Press.

Hagan, R.L. (1997). In praise of the null hypothesis statistical test. *American Psychologist, 52*, 15–24.

Harré, R. (1972). The philosophies of science: An introductory survey. Oxford: Oxford University Press.

Heidegger, M. (1996). *Being and time* [J. Stambaugh, Trans.]. Albany, New York: State University of New York Press. (Originally published in 1953)

Koch, S. (1961). Psychological science versus the science–humanism antinomy: Intimations of a significant science of man. *American Psychologist, 16*, 629–639.

Kvale, S. (1994). Ten standard objections to qualitative research interviews. *Journal of Phenomenological Psychology, 25*, 147–173.

Lakatos, I., and Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press.

Lamiell, J.T. (1987). *The psychology of personality: An epistemological inquiry*. New York: Columbia University Press.

Lamiell, J.T. (1995). Rethinking the role of quantitative methods in psychology. In J.A. Smith, R. Harré, and L. van Langenhove (Eds.), *Rethinking methods in psychology* (pp. 143–161). London: Sage.

Lamiell, J.T. (1997). Individuals and the differences between them. In R. Hogan, J. Johnson, and S. Briggs (Eds.), *Handbook of personality psychology* (pp. 117–141). Orlando, Florida: Academic Press.

Lamiell, J.T. (1998). "Nomothetic" and "idiographic": Contrasting Windelband's understanding with contemporary usage. *Theory and Psychology, 8*, 23–38.

Lamiell, J.T., and Durbeck, P.K. (1987). Whence cognitive protoypes in impression formation? Some empirical evidence for dialectical reasoning as a generative process. *The Journal of Mind and Behavior, 8*, 223–244.

Loewenstein, G., and Schkade, D. (1998). Wouldn't it be nice? Predicting future feelings. In D. Kahneman, E. Diener, and N. Schwarz (Eds.), *Understanding well-being: Scientific perspectives on enjoyment and suffering* (pp. 85–105). New York: Russell Sage Foundation.

Lykken, D. (1968). Statistical significance in psychologica research. *Psychological Bulletin, 70*, 151–159.

Lykken, D.T. (1991). What's wrong with psychology anyway? In D. Cichetti and W. M. Grove (Eds.), *Thinking clearly about psychology, Volume 1* (pp. 3–39). Minneapolis: University of Minnesota Press.

McLeod, J. (2000). The contribution of qualitative research to evidence-based counselling and psychotherapy. In N. Rowland and S. Goss (Eds.), *Evidence-based counselling and psychological therapies: Research and applications* (pp. 112–126). London: Routledge.

Meehl, P.E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*, 103–115.

Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834.

Meehl, P.E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Pscyhological Inquiry, 1*, 108–141.

Meehl, P.E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66*, 195–244.

Nisbett, R., and Schachter, S. (1966). Cognitive manipulation of pain. *Journal of Experimental Social Psychology, 2*, 227–236.

Nisbett, R., and Wilson, T. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Nunnally, J.C., and Bernstein, I. (1994). *Psychometric theory*. New York: McGraw–Hill.

Parrott, W.G. (1993). On the scientific study of angry organisms. In R. Wyer and T. Srull (Eds.), *Perspectives on anger and emotion* (pp. 167–177). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Popper, K.R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Porter, T.M. (1986). *The rise of statistical thinking, 1820-1900.* Princeton, New Jersey: Princeton University Press.
Robinson, D.N. (1995). *An intellectual history of psychology.* Madison, Wisconsin: University of Wisconsin Press.
Robinson, D.N. (1998, August). *Theory and the data it vindicates.* Paper presented at the Annual Convention of the American Psychological Association, San Francisco, California.
Robinson, D.N. (in press). Paradigms and "the myth of framework": How science progresses. *Theory and Psychology.*
Rychlak, J.F. (1981). *A philosophy of science for personality theory.* Malabar, Florida: Krieger.
Rychlak, J.F. (1988). *The psychology of rigorous humanism.* New York: New York University Press.
Schlagel, R.H. (1995). *From myth to modern mind: A study of the origins and growth of scientific thought.* New York: Peter Lang.
Schmitt, D., Couden, A., and Baker, M. (2001). The effects of sex and temporal context on feelings of romantic desire: An experimental evaluation of sexual strategies theory. *Personality and Social Psychology Bulletin, 27,* 833–847.
Serlin, R.C., and Lapsley, D.K. (1985). Rationality in psychological research: The good enough principle. *American Psychologist, 40,* 73–83.
Shrauger, J., Ram, D., Greninger, S., and Mariano, E. (1996). Accuracy of self-predictions versus judgments by knowledgeable others. *Personality and Social Psychology Bulletin, 22,* 1229–1243.
Slife, B., and Gantt, E. (1999). Methodological pluralism: A framework for psychotherapy research. *Journal of Clinical Psychology, 55,* 1453–1465.
Slife, B., and Williams, R. (1995). *What's behind the research?: Discovering hidden assumptions in the behavioral sciences.* Thousand Oaks, California: Sage.
Sullivan, H. (1952). *An introduction to the philsophy of natural and mathematical sciences.* New York: Vintage Press.
Thoresen, C. (1999). Spirituality and health: Is there a relationship? *Journal of Health Psychology, 4,* 291–300.
Valsiner, J. (Ed.). (1986). *The individual subject and scientific psychology.* New York: Plenum Press.
Wertz, F. (1999). Multiple methods in psychology: Epistemological grounding and the possibility of unity. *Journal of Theoretical and Philosophical Psychology, 19,* 131–166.
Williams, R.N. (1999, April). *On methods, methodologies, and the mythos of science.* Paper presented at the Annual Convention of the Southwestern Psychological Association, Albuquerque, New Mexico.
Wilson, T. (1985). Strangers to ourselves: The origins and accuracy of beliefs about one's own mental states. In J. Harvey and G. Weary (Eds.), *Attribution in contemporary psychology* (pp. 9–36). New York: Academic Press.
Wilson, T., Dunn, D., Kraft, D., and Lisle, D. (1989). Introspection, attitude change, and attitude–behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Volume 22, pp. 287–343). San Diego, California: Academic Press.
Wilson, T., Laser, P., and Stone, J. (1982). Judging the predictors of one's mood: Accuracy and the use of shared theories. *Journal of Experimental Social Psychology, 18,* 537–556.
Worrall, J., and Currie, G. (Eds.). (1978). *Imre Lakatos: Philosophical papers. Volume. 1: The methodolgy of scientific research programmes.* New York: Cambridge University Press.
Yanchar, S. (1997). William James and the challenge of methodological pluralism. *The Journal of Mind and Behavior, 18,* 425–442.