Psychological Reports, 1993, 72, 707–733.
Reprinted in N.G. Waller, L.J. Yonce, W.M. Grove, D. Faust, & M.F. Lenzenweger, (Eds.), *A Paul Meehl Reader: Essays on the practice of scientific psychology* (pp. 411-431). Mahwah, NJ: Erlbaum, 2006.

#157

Philosophy of Science: Help or Hindrance?[1]

Paul E. Meehl

University of Minnesota

Abstract.—Scientists routinely employ metatheoretic principles, explicit discussion of which typically occurs in times of intense controversy, theoretical crisis, scientific revolution, and entry into a new domain. The writings of philosophers, including their disagreements, are often helpful in such circumstances. Whether knowledge of formal metatheory helps us in doing "normal science" is researchable. Much scientific thinking is of poor quality, and it could be improved by explicit metatheoretical education. Clinical practice and training programs should emphasize rational skepticism, respect for evidence, objectivity, and quantitative thinking. Because the relation between principles and success is probabilistic, metatheoretical research should implement the case study method by formal actuarial procedures.

The subtitle of this article may suggest that working scientists, whether chemists or psychologists, can avoid philosophizing; but we cannot. Scientists philosophize willy-nilly, whether or not they label it as such. Consider a typical article in an experimental journal. After an introductory section describing the scientific problem, a "methods" section describes what was done; a "results" section reports observations and statistical or graphical summaries of them; mathematical derivations may be presented. The *terms* in these sections designate objects, properties, and events of the subject-matter domain. They speak of rats pressing levers, schizophrenics sorting MMPI items, college students reporting perceptions, and theoretical entities employed to "explain" these things. The logician calls such expressions *object language* because they are about the objects—observed or inferred—of which the science treats. Then comes the "discussion" section, where a different sort of statement appears and often predominates. Here, the scientist speaks *about* the object-language statements, their *meaning* and their *logical relations*, especially the *probative* relations between the theoretical and observational statements. Logicians call this *metalanguage*, since its referents are not the physical objects of the domain but the scientist's statements. The tipoff to metalinguistic discourse is the recurrence of such terms as 'prove,' 'infer,' 'valid,' 'fallacious,' 'follows from,' 'is consistent with,' 'casts doubt on,' 'assumes,' 'evidence for,' 'contradicts,' 'denotes,' 'defines,' 'true,' and 'false.' The term 'probable' is object-linguistic when it denotes a statistical relative frequency, but it is metalinguistic when it characterizes the relation of evidentiary support. The

[1] Invited address delivered at the Fourth Annual Convention of the American Psychological Society, San Diego, California, June 20-22, 1992.

metalanguage terms and statements are the philosopher's game and, in a sense, the scientist plays that game in any article containing a discussion section.

In a sense, but not quite. The scientist's metatalk *subsumes* under metarubrics and *applies* metaprinciples but does not (ordinarily) *analyze* the rubrics or *criticize* the principles as does the philosopher. The philosopher engages in meta-metatalk, he asks how a metaprinciple is to be understood and justified.

Where do these principles of scientific method come from? There is no qualitative dispute about the list of sources, only about the relative weights they deserve: generalizing current scientific practice, analyzing episodes in the history of science, formulations by reflective scientists (including the great founders of post-Galilean science, e.g., Galileo, Newton, Boyle), armchair epistemology, probability theory, and formal logic. The compelling power of these sources varies widely. For example, one distinguished scientist's pronouncements on method will often be contradicted by another's, whereas if we violate a theorem of formal logic or mathematical probability, we are in bad trouble. We may choose to ignore the philosophers on metaphysics, epistemology, or ordinary language analysis, but we cannot ignore a theorem of formal logic any more than a valid proof in geometry or algebra. *Example*: The hypothetical syllogism with major premise "If p, then q," propositions p and q being connected by the logician's horseshoe, has four figures. The first—if p then q, p, infer q—is formally valid and is called *modus ponens*. The fourth—if p then q, not q, infer not p—is also valid and is called *modus tollens*. The third—if p then q, q, infer p—is invalid, the fallacy called "affirming the consequent." But this invalid third figure is, alas, the usual form of inference in empirical science, where we reason "Theory T implies fact F; we observe F in the lab; so we infer T." This irksome situation is why empirical science is inherently probabilistic, why we can be objectively incorrect despite proceeding properly (see Faust & Meehl, 1992). Inference backward from fact to theory is not deductive, hence it cannot be "for sure." Logicians say that all induction, whether generalization of observed relations or inference to theoretical entities, is *ampliative*, that is, it goes beyond what the particulars literally assert. Even "all swans are white"—no theoretical constructions involved—obviously says something about every swan in the universe, past, present, and future, observed by someone or not observed. That strong universal statement is not deducible from any finite set of observation statements. Hence, if we do properly assert it on that basis, we can only do so in probability. The paradox that empirical research seems to rest on a fallacy led to Morris Cohen's witticism: "All logic texts are divided into two parts. In the first part, on deductive inference, the fallacies are explained. In the second part, on inductive inference, they are committed." This is no hairsplitting logician's nicety. It generates deep problems in the appraisal of theories and is the rationale of Sir Karl Popper's emphasis on falsification (see Popper, 1959, 1962, 1983). Indirectly, it underlies the criticisms of null hypothesis testing by myself and others (Meehl, 1978, 1990a, 1990d; Morrison & Henkel, 1970).

The biggest change in philosophy of science since I was a student is the emphasis on history of science, the data base for a discipline conceived as itself empirical. I favor this change, but I disagree with some of the philosophical conclusions drawn from it. I also hold that the new orientation is not being properly implemented *on its own terms*.

The new emphasis is linked with Thomas Kuhn's (1970) book on scientific revolutions and the subsequent orgy of "positivist-bashing." Despite the merits of his book, Kuhn has not convinced most historians, philosophers, or scientists of his main theses. That does not prove he is wrong, but it does mean that psychologists lacking expertise in these matters ought not to brandish Kuhn's book at their opponents, rather like the TV evangelists wave the Bible to settle everything. For example, I often hear and read the flat statement, "As Kuhn has shown, all observation is theory laden." Kuhn *has not shown* that, he has merely *asserted* it; and only a minority of philosophers and scientists agree with him. The alleged theory-ladenness can be made plausible for quantum mechanics and relativity theory, which are spooky and not representative of other empirical sciences. Experimental nuclear physics is special because its "observational devices" include *other* particles than the ones being studied, so the interpretation of a macro-instrument reading relies on the theory of particle interactions. But "theory-laden" is too vague a formulation, and the portions T_1, T_2 of theory T relied on in an experiment appraising *another* portion T_3 of theory T may have been independently tested by experiments *not* relying on T_3.

Consider the protocol: "Rat 13 locomoted to the right side goal box and there drank some water." What is theory laden about that? Nothing. But isn't the term 'goal' theory laden? It looks so, but it is not. In Tolman's *theory*, it is where the rat expects to find water. In Hull's, it is where a drive reduction occurs; in Skinner's, it is the locus of a reinforcing event. But in the neutral *observation language, which they share*, it is the box at the right end of the T cross-bar where rat 13 has encountered water and ingested it. "The subject, after being instructed to look at a red circle for 30 seconds and then look at the wall, reported 'I see a large washed-out blue-green circle'." What is theory laden about that? Nothing. You may be thinking, "Well, Meehl makes it easy for himself by picking simple cases from animal behavior and human perception. But as a clinician, he should know better." Not so, and I am a 33% Freudian (used to be 67%). Take a clinical example: "In Alabama, students with a 9′4 MMPI profile are twice as likely to ignore a letter from the college dean as are those with a 7′2 profile" (cf. Griffith & Fowler, 1960). Theory-laden? No. Or, behind the couch: "After reporting a dream of squirting a garden hose on a fire, this (male) patient spoke the words 'shame,' 'proud,' 'ambition,' and 'embarrass'." Again, nothing theory-laden at all. If the analyst had said, "Most of the session material concerned the topic of ambition, as shown by words thematically linked to that motive," we would have a low-order theory-ladenness. Even that can be objectified by advance stipulations to tally entries in Roget's *Thesaurus* or empirical word association norms (e.g., Palermo & Jenkins, 1964), but based on motivational/affective stimulus words. Furthermore, the auxiliary minitheory relied on is in semantics and psycholinguistics; it does not in the least depend on the theory being tested, Freud's urethral triad. What we *observe* in the psychoanalytic hour is speech (words, rate, pitch, timbre, volume, pauses), posture, gesture, facial expression. Everything else is theoretical inference (Meehl, 1983).

Psychologists with obscurantist motives for invoking Kuhn tend to conflate several issues (e.g., valid role of clinical intuition, use of empathy and recipathy, "foundationalism" in the early positivists, dogmatic attitudes of some behaviorists, unavoidability of open concepts) with a basic issue concerning *all* empirical knowledge claims, inescapable in *any* domain ("scientific" or not), namely, "Does this knowledge claim come with trustworthy credentials?" When that question is pressed by a fair but skeptical interlocutor, no quotes from Kuhn can avoid the task of *proving* what one claims to have observed, and in a way that does not require the skeptic to accept one's theory, *that being what is in dispute*. I am sure that Thomas Kuhn never meant that scientists should feel free to commit the *petitio principii* fallacy.

A beautiful example of what actually happens when two theorists' observations are "theory laden" (in the bad sense) occurs in the debates of the British Psychoanalytic Society concerning the views of Melanie Klein:

Dr Payne: I should like to ask Miss Freud a question. She said she did not think that the mother was regarded by the child as an object in the first year of life. (Miss Freud: First half year.) There are behaviouristic facts, for instance, that the child may be given the bottle by a stranger and refuse it, just as it may be given a wet nurse and would not take the breast from her. Such facts suggest that the child differentiates between objects.

Dr Glover: Would Miss Freud like to answer this question now?

Miss Freud: Yes, this is a question that should certainly not be answered too quickly or without preparation, but I would not have made the remarks I made without considering such facts. There is no doubt that the child notices differences and even the slightest differences in atmosphere, position, behaviour, mood of the person who gives the bottle. But, as I said before, these differences reach the child because there is a change of the form in which the satisfaction is presented. It is not the same as the beginning object relationship we find in the second half year where the child will refuse just because it is a different person. What we call a change of atmosphere, change of mood, it is not the tie to the person. If one could imitate the form in which the satisfaction is given completely, the child would accept it. Experience shows that it is surprisingly easy to interchange objects at that time, just as it is surprising how difficult it is later.

Dr Payne: It seems to me the difference is one of degree.

Miss Freud: It seems so absolutely different in kind when one watches it closely.

Dr Payne: *But I have watched it too*.

(King & Steiner, 1991, pp. 435-436; italics added)

After Dr. Payne's blunt, "But I have watched it too," what can be said? Without a proper sampling with sound movies, nothing. This dialectic situation is the kind in which philosopher Alburey Castell (the "Augustine Cassell" of Skinner's novel *Walden Two*) used to say, "Full stop." This sort of theory-ladenness is not a Kuhnian virtue, it is a scientific vice. A refreshing contrast, involving a similar task, is the fascinating finding that home movies of children reveal gesture patterns that predict which sibling will develop schizophrenia as an adult (Walker & Lewine, 1990). Efron and Foley (1937) objectively analyzed gestural patterns of Italian and Jewish immigrants by drawing lines of motion of anatomical reference points (e.g., shoulder, wrist, elbow, chin) on consecutive movie frames. For proof of such relationships in the analytic session, see Mahl (1987). When disagreement persists as to what observable events took place, there can be no substitute for *objective recording* and *objective classifying or measuring*. If the latter cannot be achieved by physical specification, open to any attentive observer with normal sensory function, but requires a specially trained observer, then the "skilled judgment" must be *uncontaminated* (e.g., in the Walker and Lewine study the raters did not know which child would become schizophrenic 20 years after the movie was made). These elementary points about how scientific knowledge claims differ from folklore and superstition should be too obvious to require explanation; but, alas, some current obscurantist writing makes it necessary. One cannot validly move from "1928 Vienna logical positivism contained some errors" to "Kuhn has liberated us, so now we can say anything we please and bear no responsibility for the correctness of our observations."

The confusion wrought by Kuhn (or misreadings of him?) was cleared up 20 years ago by Carl Kordig (1971), using Kuhn's own example of 'the sun rises in the East.' For a recent powerful critique of Kuhn's reliance on psychological research, see the excellent article by Gilman (1992). Theory motivates one to enter the lab or the clinic files, and it suggests what to control or manipulate, what aspects of behavior to record (although if it blinds you to everything else that occurs, you are a poor scientist), perhaps the metric to employ, and the statistical analyses. Who has ever denied any of this? Certainly not the logical positivists, who usually didn't bother to mention these matters because they are so obvious and trivial. What theory is *not* allowed to infect is, of course, the *content* of the observation—the meter reading, the color, the phonemes uttered, the pH, the cumulative record, the Pearson correlation between scores x and y. If *that* is theory laden, we do not call you a liberated Kuhnian, we call you a fraud, a data-faker, a scientific crook.

Much of the positivist-bashing distorts what the Vienna Circle said[2] or takes a single mistake by a single member—even if corrected shortly thereafter—as a group dogma. The worst mistake was Schlick's, "The meaning of a sentence is the method of its verification," held briefly by some of the group. Similar errors were made by a few Americans, but under the influence of Bridgman's operationism, quite independent of Vienna. Herbert Feigl was the one who in 1924 first urged leader Schlick to hold regular, organized meetings. He co-authored the first paper in English on the Vienna position (Blumberg & Feigl, 1931) five years before Ayer's (1936) bombshell book *Language, Truth, and Logic*. He invented the term 'logical positivism' and was the leading expositor of the doctrine in the United States. He came to Minnesota in 1940 and was my teacher, then co-author and co-instructor, and personal friend for 30 years. I know what the positivists believed, where they disagreed among themselves, and how their views evolved. Much of what is attributed to them by the positivist-bashers is simply incorrect. When psychologists in the "soft" areas write about their liberation by Kuhn and Co., I want to inquire of them "What was the ingenious experiment you devised but never performed because of the wicked positivists? What is the exciting theory you conceived but feared to publish lest the positivists make fun of you?" I have done theoretical and empirical work in several areas of psychology, including latent learning, MMPI scale construction and profile analysis, psychometric theory, clinical prediction, interview assessment, political behavior, forensic psychology, theory of schizophrenia, behavior genetics, and taxometric methods. All of these topics present complex conceptual problems involving what Arthur Pap (1953, 1958, Chapter 11, 1962, Chapter 3) called 'open concepts,' for which any strict operational definition is a fake. But I have never felt the

[2] A pamphlet, *Wissenschaftliche Weltauffassung: Der Wiener Kreis*, a collaborative effort by Carnap, Hahn, and Neurath, aided by Feigl and Waisman, was published in 1929. The "orthodox" Vienna Circle position is succinctly presented in that famous manifesto (of which the Circle's leader Moritz Schlick, absent in the U.S. when others wrote it, disapproved). For a concise overview, see Passmore's (1967) entry in the *Encyclopedia of Philosophy*. An early classic critique of the original position is Weinberg (1936). A nice collection of essays was edited by Ayer (1959). There have been hundreds of books and thousands of articles devoted to the various aspects of the Logical Positivist position and its amended form, usually called 'Logical Empiricism.' Reichenbach (1938) wrote a beautiful exposition, still very worth reading whether or not one agrees with him about induction and probability. The increasingly "tolerant" variant (evolved as Logical Positivism → Logical Empiricism → Empirical Realism) is presented clearly, forcefully, yet with scrupulous fairness in Feigl's collected essays (edited by Cohen, 1981; see especially Feigl, 1969). Any treatise on philosophy of science (such as those cited in the bibliographic note at the end of this article) unavoidably discusses the movement, although some of the more recent discussions are unfair and careless. Contemporary psychologists grinding an obscurantist axe often indulge in caricature—*caveat lector*!

least bit constrained, cramped, or inhibited by the logical positivists or their intellectual descendants. On the contrary, my thinking about these deep and difficult matters has been facilitated by my reading and discussion with such first-class intellects as Feigl, Carnap, and Hempel. I say the same for Reichenbach who in his classic 1938 *Experience and Prediction* uses the term 'positivists' to denote his Viennese opponents—hence the positivist-bashers who label him a positivist thereby reveal that they have not even read his book! I have also learned much from anti-positivist Popper and his erstwhile disciples Feyerabend and Lakatos. The way to treat philosophers of science is to pick and choose, take what you can use, and in the process acquire philosophical sophistication from their disputes.

Modification of the original doctrine was largely due to critical analysis by the positivists themselves or by others sympathetic to and in close contact with them (e.g., Nagel, Pap, Popper, Reichenbach). One property of respectable sciences is that they do evolve, reject some conjectures and amend others, and have a cumulative character. Other than symbolic logic, logical empiricism is the only philosophical movement that is similar to science in this respect. Furthermore, as Popper emphasizes, a clearly falsified theory is not a "failure," as we can learn from our mistakes. Unraveling a deep, subtle, complex mistake, whether in a scientific theory or in metatheory, is always illuminating. In the years since the logical positivists' 1929 manifesto (see Footnote 2), philosophers of science have definitively solved some problems, rejected proposed solutions to some, and clarified the formulation of others. We can learn from all three categories. We do not dismiss a scientific theory because it is not yet "settled"; why should we impose that perfectionist demand on the philosopher? Here is a list of some metatheoretical topics which, regardless of their state of acceptance, are helpful in appraising theories:[3]

Orders of dispositions: A first-order disposition of an iron bar might be that it is *magnetic* (i.e., its tendency to move toward a large mass of iron or its tendency to attract nonmagnetized iron objects such as a nail or a pin; some prefer to call the latter a 'power' rather than a 'disposition' because it is active rather than passive). A piece of iron can be magnetized and can also be demagnetized; therefore, *magnetizability* is a disposition of the second order. If a nonmagnetizable element can be transmuted into a magnetizable substance, that would be a third-order disposition, and so on. Metaphysically speaking, that an entity or substance is, so to speak, "what it is rather than some other kind of entity or substance" is given by its

[3] Readers of this chapter in draft urged me to do more than simply list these topics. A complete and fair explanation of any one of them would take a great deal more space, but I have tried to give enough explanation for interested readers to get started if they wish to pursue the topics in more detail.

supreme dispositions (Broad, 1933). In psychology, a habit or atomic trait is a first-order disposition, the ability to acquire such a habit or trait is a second-order disposition. Achievement in a domain (e.g., spelling, English composition, mechanical facility) is a collection of related first-order dispositions. The ability to acquire such achievement classes is a second-order disposition (measured, say, by an intelligence test or a mechanical aptitude test). The capacity of the central nervous system of an individual to develop such abilities (if not deprived of food, oxygen, etc.) is a third-order disposition. One might think of the heritable component of *g* as a fourth-order disposition. The relation of traits in differential psychology and personology to genetics and learning theory concepts cannot be properly analyzed without understanding orders of dispositions (see, e.g., Meehl, 1986).

Surplus meaning of theoretical terms: Theoretical terms in all sciences, from physics to the social sciences, have a meaning over and above the observational properties or dimensions that we employ in inferring them. Even simple dispositional terms (e.g., soluble, flammable, frangible, irritable) in which it looks as if no theoretical content is involved have a slight surplus meaning from the strict epistemological standpoint, as has been pointed out by Popper and others, because even if they do not invoke unobservable entities (like protons or libido or genes), they do, at least implicitly, predict other observations that have not in fact been made when we attribute them, observations which it is always possible might surprise us if sampled under suitable circumstances.

Classification of entities: A division of basic kinds of entities I have found useful in analyzing psychological theories is six-fold, not prejudging the possibility that some may be reducible to others: substance, structure, event, state, disposition, field.

Nomothetic and idiographic explanation: The former is explanation of a particular event in terms of general laws, as we do in astronomy or chemistry; the latter is more like a narrative concerning a single event, as in understanding a life history or a historical process such as the fall of Rome (Allport, 1937). Whether idiographic explanation is always in principle reducible to nomothetic continues to be in dispute among methodologists.

Convergence of evidence versus prediction of novel facts: It seems generally held (but not universally) by practicing scientists as well as by philosophers of science that, other things being equal (e.g., the logical structure of the inferences used), after-the-fact explanation does not provide as strong corroboration of a hypothesis or theory as does the correct prediction of a novel fact.

Role of auxiliary theories in falsification of a particular theory: Since the derivation of a prediction from a theory always involves some auxiliary theories as well as a *ceteris paribus* ("all other things being equal") clause, the formal logic of an experimental test is not, strictly speaking, a direct test of the theory alone; it is also a test of all the auxiliaries needed (Meehl, 1978, 1990a, 1990b).

Strategy and rationale of Lakatosian defense against falsification: An apparent falsifier of the theory may be avoided by challenging the other assumptions or by adding an additional entity or theory, thus challenging the *ceteris paribus* clause. When this is a legitimate move and how long it may be continued to defend a theory against a series of apparent falsifiers is at present imperfectly worked out (Meehl, 1978, 1990a, 1990b).

Appraising progressive versus degenerating research programs: When adjustments to ward off apparent falsifications become excessive, the research program is "degenerating." Criteria are needed for deciding when this is taking place (Lakatos, 1970).

The so-called law of parsimony: Despite its widespread invocation by scientists when it suits their purposes, the status of the law of parsimony is not clear, and treating it as some sort of mandatory legislation that says one must always prefer the "simpler" explanation of everything has not been philosophically defensible. No satisfactory general definition of 'simplicity' has been invented. Computational simplicity in a technological context is obviously desirable, but not of great theoretical interest. There is no epistemological or metaphysical basis for saying that we know in advance that the world has to be simple; some of the most powerful developments in science are obviously not simple at all. The tensor calculus and the number of dimensions involved in General Relativity are far more complicated than Newtonian theory, but Einstein appears to be right and Newton wrong. One kind of simplicity is advocated by Popper as being related to falsifiability; hence, simpler (more easily falsifiable) hypotheses are preferable to less easily falsified complex hypotheses. Finally, there is Ockham's Razor, which amounts essentially to saying that one ought not to invent a theory to explain something when we already have a well corroborated theory capable of explaining it.

The Ramsey Sentence: This is a technical device of symbolic logic permitting the elimination of theoretical terms without liquidation of the theory (see, e.g., Carnap, 1966, Chapter 26 and pp. 269-272; Lewis, 1970; Maxwell, 1962, pp. 15ff, 1970, pp. 187-192; Watkins, 1984, pp. 146-148; for criticism see Glymour, 1980, pp. 20-29). Its clarifying value for us is that it shows rigorously how expressions can both define concepts and assert facts concurrently, contrary to the simplistic operationism still being taught in some introductory psychology classes.

Distinction and relation between interpretive and operational text: A mathematical formalism (e.g., factor analysis, taxometrics, learning theory equations) becomes empirical by being associated with an embedding text of words or diagrams that designate physical entities. There is an important distinction between the "operational text," which is what coordinates a proper subset of the theoretical terms to observational terms, and the "interpretive text," which does not perform such a direct linkage to the data language but which characterizes the theoretical entities in various ways *in the language of the theory itself*. Whether all the terms implicitly defined in a nomological network can be given adequate operational explication by what I call "upward seepage" from the data (via the Ramsey Sentence) is unclear, but I very much doubt it (Meehl, 1978, 1990a, 1990b). For that reason we should maintain the distinction between embedding text which is operational and embedding text which is not.

Testability versus confirmability: A disposition or more complicated theoretical attribution is testable if there is an operation we can perform, and a resulting observation we can make, that will corroborate or refute the attribution. But it may happen that we cannot (for either technical or ethical reasons or perhaps because of incomplete theoretical understanding, e.g., auxiliary theories) perform an operation, although we can "wait around," hoping to see something happen which, if it does happen, will confirm or refute the theory. Confirmability only requires that we be able to notice something if it does happen, it does not require that we be able to perform an operation that makes something happen. Carnap (1936–37) opted for confirmability as a more tolerant criterion and gives examples of why testability is too strict for scientific purposes. "Person X is hysteroid" may not be testable—we cannot at will precipitate a hysterical conversion—but if that person does subsequently develop a paralyzed arm under stress (Adolph Meyer's "experiment of nature"), the hysteroid disposition stands corroborated.

Severity of tests of theories: Theories which make very broad, loose, or tolerant predictions are corroborated if they succeed, but not as strongly as they would be had they made very risky, narrow predictions (e.g., numerical point predictions, or predictions that fall in restricted intervals that occupy only a small region of the antecedently possible numerical values, or strange, counter-intuitive qualitative effects). While severity of tests is associated with the views of Sir Karl Popper, inductivists like Wesley Salmon have a place for it also from a different perspective, one in which Bayes' Theorem plays a part and the successful prediction would constitute a "damn strange coincidence" if the theory had no verisimilitude. It is obvious from the history of science that working scientists put a great weight upon this matter of riskiness or strange coincidence even if they do not engage in formal philosophy of science (Meehl, 1978, 1990a, 1990e).

Asymmetry of confirmation and falsification: This is a matter of formal logic. A falsifying fact, if accepted and without fiddling with the auxiliaries, logically entails that the theory is incorrect, whereas a confirming fact (being in an invalid figure of the syllogism) gives it "support" but does not, strictly speaking, entail it.

Three kinds of ad-hoc-*ery*: A falsifying fact may be warded off by introducing an auxiliary hypothesis or theory that does not predict (or forbid) anything new. Or an auxiliary hypothesis or theory may be introduced that makes new predictions, but the predictions it makes are not empirically successful. Or a new auxiliary hypothesis or theory may add content by predicting new observations, and it may also have some empirical success, but it is simply a "local patch job" in a ramifying nomological network that lacks the unity characteristic of a coherent research program or a powerful theory sequence (Lakatos, 1970). For example, if Skinner were to explain latent learning in the operant chamber by invoking Freud's speculations in *Beyond the Pleasure Principle*, that would be the third kind of *ad hoc*, even if it "worked" predictively.

Reduction of concepts and reduction of laws: The laws at a given level (e.g., molar behavior) cannot be explained from laws at another level of analysis (e.g., how the brain works) unless the molar concepts are reducible to the molecular ones. Therefore, reducibility of concepts is a necessary condition for reducibility of laws, but it is not a sufficient condition, as is shown by the perennial controversies about emergence.

Paradoxes of confirmation: Hempel and others have pointed out oddities about confirmation syllogisms, such as that 'all crows are black' is loosely thought of as being confirmed by accumulating particulars (i.e., a black crow, another black crow, another black crow, and so on). However, the statement 'no ~~non-white~~ [non-black] birds are crows' is logically equivalent to 'all crows are black.' Hence, observing a white swan, which is a nonblack noncrow, is a confirming instance of 'all crows are black'; and if we accumulate a large number of white swans, green parrots, blue pigeons, and so on, we are piling up instance confirmations of the original generalization that all crows are black. Obviously science does not work that way, so something is wrong in the logical model. Whether the various analyses of this and related paradoxes have taken care of the problem, I do not know, but there may be kinds of "clinical experience evidence" that are of similar form to Hempel's paradoxes and give psychologists an illusion of strong evidentiary support.

Open concepts: Except for strictly operational definitions (which turn out to be extremely rare, some would say nonexistent even in physics), theoretical concepts are "defined" implicitly or contextually by their role in the theoretical network, which latter is empirical by the fact that a proper subset of the statements that occur as derivable theorems in the network are fairly directly coordinated to observational statements "confirmed" in the laboratory, by statistics in the clinic files, or by naturalistic observations in the field (see, for psychology, Cronbach & Meehl, 1955). Because the list of indicators of a theoretical concept is extensible and usually is, in fact, enlarged by further knowledge, the concept is to that extent flexible, modifiable, "open." Further, even when *strands* in the nomological network (laws connecting *nodes*, which are theoretical entities related by those laws) are identified, in the life sciences and very much in the behavioral sciences, these strands themselves are probabilistic connections rather than strict nomologicals; so that even if one had a network considered to be complete as regards the connections and the entities, there would still be a certain stochastic looseness about the concepts (e.g., as when we deal with a gene of incomplete penetrance, or recognize measurement errors in intelligence testing). Thus, even if the theory were quasi-complete, the application of the implicitly defined theoretical concept to an individual would still be only probabilistic. Finally, a third kind of openness (see Meehl, 1977) consists of the fact that, in addition to knowing the role it plays in the network, its dispositional properties and relations to other things, we want to know the inner nature, the structure or composition, of a theoretical entity. The process of filling in the nodes of the nomological net is an important part of the research program of any science. Super-operational psychologists are sometimes phobic about queries concerning the "inner nature" of a theoretical entity implicitly defined via the nomological network, but we need not be. Physicists, chemists, physiologists ask and answer such *compositional-structural* questions all the time. The greatest scientific triumph of our half-century was Watson and Crick's on the nature of the gene—a concept originally defined by its role in the statistical network of phenotypic breeding relations.

The semantic conception of truth: Many social scientists have a phobia about the word 'true,' because they mix up its ontological meaning with the evidentiary question of certainty. 'Caesar crossed the Rubicon' is true only if he did so, and that one historical event is the simple, sole *truth condition* for the sentence. But of course the *evidence* for that sentence may be quite inconclusive, and even if apparently conclusive, might turn out to be erroneous. Even if we humans never discover it is wrong, it would nevertheless be wrong if Caesar did not in fact cross the Rubicon. My evidence is some mounds of ink on a palimpsest in the Vatican library, whereas the centurion who rode across on horseback beside Caesar had very different evidence. Yet, the *content* of the statement, whether spoken in Latin or English, is the same. (This suffices to refute the notion that the meaning of the sentence consists of the method of its verification!) There is no reason why psychologists should be phobic about the term 'truth'; it is simply a predicate of the metalanguage, and the conditions for appropriately uttering it are straightforward. The evidence that warrants us saying "Caesar crossed the Rubicon" is precisely the same as the evidence that warrants us saying in the metalanguage, "The statement 'Caesar crossed the Rubicon' is true."

Probability as a physical relative frequency versus as evidentiary support: On the face of it, there seem to be two qualitatively distinct meanings of the word 'probability,' as set out by Carnap (1945). He uses the term 'probability$_1$' to designate evidentiary support or degree of confirmation (as when we say there is a lot of evidence to support Einstein's theory, or the theory of the gene, or Darwin's theory), distinguishing it from 'probability$_2$', a relative frequency of an attribute or event in a specified class of physical entities or occasions. Disagreement persists as to whether it is possible to reduce all probability concepts to relative frequency, but no logician has yet been able to do that, even if it is possible in principle. Unfortunately, statistics textbooks rarely mention this distinction, and, an associated omission, they rarely mention the important distinction between a *statistical hypothesis* concerning some parameter of a physical population and a *substantive causal theory*. Hence, the student, unless more philosophically sophisticated than most social science majors, comes to more or less equate substantive theory with statistical hypothesis and to equate the appraisal of substantive theories with something like, or even identical with, the testing of a statistical hypothesis, which in turn has led to the widespread abuse of null hypothesis refutation as a powerful way of corroborating substantive theories, which it is usually not (see Meehl, 1990a).

Causal laws and accidental universals: "Stress is proportional to strain" (Hooke's Law) or "acceleration of a mass is proportional to the applied force" (Newton's Law of Mechanics) appear to have the same logical form as "all the coins in Meehl's pocket are silver" or "there is a statistically significant negative correlation between college grades and hours of study." Yet the first two are laws of nature and they warrant counterfactual conditionals, whereas the second two are not laws of nature and do not warrant counterfactual conditionals. That is, from the third statement one cannot infer, "If this penny *were* in Meehl's pocket, it *would* be silver"; nor can one rely on the fourth statement to advise a student who is getting poor grades to study less in order to creep up on the regression line toward higher grades! It has been pointed out by Carnap and others that the "laws" of biology, being structure-dependent, could have been otherwise had evolution taken a different course on our earth, so that (strictly speaking from the standpoint of the logician) the laws of biology are also "accidental universals" rather than laws of nature. Interpreting partial correlations in analyses of covariance and path coefficients in the social sciences is a tricky business, partly because of this distinction.

Counterfactual conditionals: The proper logical analysis of conditional statements contrary to fact (e.g., if this piece of butter had been heated, it would have melted) has been a problem for the logicians for many years and is still not completely solved, but some sophistication about the moves in that game is important for the interpretation of counterfactuals in psychology.

Explicit and implicit definition: Most scientific theoretical terms are not explicitly defined according to the criteria of Aristotle's Five Rules of Definition or of the rules for definition one learns in an English course. The most powerful theoretical constructs in the advanced sciences are defined implicitly by their roles in the theoretical law network, and even those that can be explicitly defined are defined explicitly using other theoretical terms that themselves are only defined implicitly. Thus, for example, in physics we may define 'temperature'—the theoretical construct, not the thermometer reading, which is a fallible indicator of it—by the usual "temperature is the mean kinetic energy in the molecules"; this is explicit, but it is not operational because, of course, the term 'molecule' is itself a theoretical term and that term is defined implicitly rather than operationally.

The bidirectional control between theory and observational protocols: This is a deep matter not yet fully explicated. In any mature science, some alleged observations are not admitted into the corpus of beliefs because a well corroborated theory precludes them. Yet, if we are empiricists, we insist that theories must bow to facts; hence, a paradox. Roughly, the solution is that *individual* candidate protocols are controlled by theory *in the short run*, but theories are controlled by the *whole collection* of admitted protocols *in the long run*.

Theoretical argument in psychopathology is permeated with philosophical mistakes. When I praised the research on family expressed emotion as a potent factor in schizophrenia relapse, a practitioner asked, "But, Meehl, how does that square with your theory of genetic etiology?" No psychologist learned in philosophy could commit such an elementary blooper. Of course, some persons just naturally think better than others, without special education in "how not to think." But I conjecture that *some* education in metatheoretic thinking can be prophylactic for all of us. In complex causal systems, it helps to be clear about just what question one is asking. The MZ twin concordance for tuberculosis was formerly in the .80s, so why do we consider it an infectious disease while we consider schizophrenia (concordance around .50) a genetic one? PKU is preventable by dietary restriction of phenylalanine, but we call it a genetic disorder. Careful analysis shows that the commonest etiological situation is that a causal factor is an *insufficient* but *necessary* component in a complex which is *unnecessary* but *sufficient* for the result. An unphilosophical psychologist could figure that out with some hard, clear-headed work. But none did. And we need not go to the trouble, because philosopher Mackie did it for us 28 years ago, in his classic paper (1965) on the INUS condition. (INUS is the acronym for insufficient-necessary-unnecessary-sufficient.) If you don't know about INUS conditions, you cannot theorize competently about concepts such as schizophrenia.

In the first decade after presenting my theory of schizophrenia, I made a collection of such metatheoretical errors, some subtle and plausible, others simply stupid. Here is a sample:

> … "You cannot study the genetics of schizophrenia until agreement exists on a *definitive set* of diagnostic signs." "To add a new symptomatic indicator to the list constituting a syndrome, or to justify a shift in the diagnostic weights within the received list, either (a) is an arbitrary redefinition or (b) requires nonsymptomatic criteria to validate it." "To rediagnose a case because its subsequent clinical course disconfirms expectation is an arbitrary act (or, 'circular')." "To say that 'true schizophrenia' refers to the genetically determined cases and all others are phenocopies is viciously circular." "We cannot assign differential diagnostic weights to the elements of a syndrome unless we have an external criterion, as in neuropathology." "Since all classifications are arbitrary anyway, and mental patients differ from normal persons in ways that exist in all degrees, it makes no scientific sense to ask whether an entity like schizophrenia 'really exists,' and the use or avoidance of this concept is a matter of preference only." "It is inadmissible to explain a given symptom as caused by a disease D unless we can define the term 'D' independently of its symptoms. Otherwise we would be mixing empirical relationships and meaning stipulations." "Any diagnostic cutting score on a continuous indicator variable will be arbitrary, a matter of semantics or convenience." "I can find you a so-called 'schizophrenic' who is more similar symptomatically to some manic-depressives than to most schizophrenics, which proves there is no such entity as schizophrenia." "To speculate that a particular person has the disposition to schizophrenia even though he has survived the morbidity risk period without becoming clinically schizophrenic is scientifically meaningless."
>
> None of these familiar remarks is expressed in technical philosophese; but they are all methodological in nature. *And they are all erroneous*. The last one, for example, imposes a criterion of empirical meaningfulness whose grave disadvantages were already shown by Carnap more than a half-century ago—when the philosophy of science was far more "operational" and "positivistic" than today. I doubt one could find a single contemporary logician or historian of science who would accept the remark quoted. (Meehl, 1972, pp. 21-22)

The current controversy over path analysis can almost be called a scandal. Invented over 70 years ago by biologist Sewall Wright (1921) to discern causal relations from the statistical pattern of nonexperimental data, it surfaced in social science in the 1960s. Some psychologists consider it the only proper way to study certain problems, while the distinguished mathematical statistician David Freedman thinks it is nearly worthless. I understand that whether your million-dollar project on drugs or delinquency gets funded depends on whether a Freedman or Bentler disciple is on the Bethesda committee! Now, there is almost nothing controversial about the mathematics, which is high school algebra. The big dispute is a *philosophical* one, involving such metatheoretical issues as inductivism versus Popperian deduction, how to test models known to be literally false, the dependence of falsification on the statistical power function, comparing two false theories' verisimilitude, numerifying predictive "riskiness," and the rationality of causal inferences from nonexperimental correlations. Neither of the parties to this dispute is formulating it correctly, chiefly because they are operating with an inadequate philosophy of science. I will not claim I have solved the problem, but I am pretty sure I know the general way to proceed (Meehl, in preparation [Meehl & Waller, 2002; Waller & Meehl, 2002]); and my constructive approach has been informed by what I learned from the philosophers.

Up to now I have spoken conservatively, as might be expected from one of my generation, a Minnesota PhD (1945) reared on amended logical empiricism. Now I shall make a radical proposal. Conceiving metatheory as the empirical theory of scientific theorizing, we know that formal logic, probability theory, and basic, unproblematic armchair epistemology do not suffice to provide principles of scientific method which aim to generalize the practices of successful science. But when we cull a list of such principles from the writings of scientists, philosophers and historians of science, we are disconcerted by the texts. Most of the principles are incommensurable, involving properties or relations that neither entail nor contradict one another. *Examples*: "We prefer theories that correctly imply numerical observational values that are precise, that take a high risk." "We prefer a theory that forecasts novel facts rather than one that only explains facts already known and used to concoct the theory." "We prefer theories that imply qualitatively diverse facts, illuminating widely different realms of experience." This independence of metapredicates produces conflict between principles when appraising theories and highlights the point that we deal with *principles, not rules* in the strict sense. The so-called "scientific method" is not a set of rules such as reading the mercury meniscus or summing degrees of freedom in an analysis of variance. Second, there is not even qualitative consensus. For example, the preference for forecasting is crucial for Popper, a mild preference for most scientists, and irrelevant for Carnap. ("But Meehl," Carnap asked me, "how could the mere date of a fact affect its logical relevance?") Since the principles can countervail each other, and none of them is a guaranteer of truth, *their relation to epistemic success is probabilistic*.[4] Now, how does one investigate probabilistic relations? By actuarial methods. Hence my thesis: "Metatheory should be conducted by random sampling of episodes in the history of science, applying formal statistical and psychometric methods to analyze the results." This Strong Actuarial Thesis I also call the Faust-Meehl thesis, since David Faust and I seem to be its only proponents (Faust, 1984; Faust & Meehl, 1992; Meehl, 1990a, 1992a, [2002, 2004]).

Metatheorists are trying to do the empirical part of their job relying on case studies. This will not work unless the episodes studied are (a) representative and (b) analyzed actuarially. Obviously a case study in which

[4] This is obvious to philosophers, but was queried by some reviewers. Suppose two theory-appraising metaprinciples P_1 and P_2 sometimes clash as to particular theories, which we know happens. But each principle has *some* validity, or we would not be using it. Unless P_1 always trumps P_2, or conversely, instances must arise for each when it gives the incorrect appraisal. Hence, at best, its epistemic correlation with truth is probabilistic only. If there were a perfect principle P_i that *always* trumps *all* others, (i.e., an infallible litmus test of theoretical truth) we would employ it only, discarding all others. But no scientist does this, and no philosopher has ever argued for any such.

Scientist X followed Principle A with success cannot answer the probabilistic question. Case studies are a source of conjectures and can sometimes *refute* a generalization if it is stated as a rule, which a metatheoretical principle is not. Most case studies refute rules that nobody maintains.

That the Strong Actuarial Thesis is propounded by two clinical psychologists is not as strange as it seems. Faust's path-breaking book, *The Limits of Scientific Reasoning* (1984), was inspired by my earlier monograph on clinical versus statistical prediction (Meehl, 1954). In 1954 I could find only a score of studies comparing the clinician's informal, impressionistic judgments with those of a regression equation or actuarial table. My colleague William Grove is now conducting a meta-analysis of approximately 175 comparative investigations, and the accuracy of the actuarial method of integrating data equals or exceeds that of the clinician in almost every study (Grove, 1986; W. M. Grove, personal communication, 1992; [Grove, Zald, Lebow, Snits, & Nelson, 2000]). I know of no controversy in social science in which the research is as massive, varied, and consistent as this (Dawes, Faust, & Meehl, 1989). The prediction tasks range from academic success to cancer survival time to winning high school football games. Elsewhere (Meehl, 1992a) I have listed 37 sources of inaccuracy in clinical assessment by the usual informal method. *All but three of these cognitive deficiencies apply to the scientist appraising theories as we customarily do*, and most of them to the metatheorist evaluating principles by case study. It is not surprising that there is marked metatheoretical disagreement, as revealed by the Talmudic compilation of Laudan, Donovan, Laudan, Barker, Brown, Leplin, Thagard, and Wykstra (1986).

The thinking of scientists, especially during controversy or theoretical crises preceding Kuhnian revolutions, is often not rigorous, deep, incisive, or even fair-minded; and it is not "objective" in the sense of interjudge reliability. Studies of resistance to scientific discovery, poor agreement in peer review, negligible impact of most published papers, retrospective interpretations of error and conflict—all suggest suboptimal cognitive performance. As to formal logic, Mahoney and Kimper (1976) have shown that 39% of physicists and 67% of biologists do not even recognize *modus tollens* as a valid syllogistic figure! I urge you to flesh out these research findings, as I have done for several years, by critically examining the reasoning in scientific articles in which you have no ego involvement. Once alerted to the problem, one finds much of the reasoning surprisingly shoddy, ranging from mild semantic slanting and puffing relevant but weak points to gross errors (illicit distribution, four terms fallacies based on equivocation) and even plain misstatements of fact. *Example*: The phrase 'proof of' has at least seven meanings, ranging from the weak "has some probative relevance," through the moderate "renders more probable than not," to the extreme "demonstrates beyond reasonable doubt" or "deductively entails."

I find that scientists who say "Fact *F* is not proof of theory *T*" almost never make clear which of the seven meanings they intend, thereby presenting the careless or naive reader with a projective test. One wonders how many budding meritorious conjectures of young unknowns have been crushed by a prestigeful figure thus abusing the phrase "is not proof of." And these seven meanings are not philosophers' hair-splitting—the term 'proof' is an important word in meta-talk, since proving and disproving is what we are about when we collect data on scientific theories. *Example*: In 1962 (see also Meehl, 1990c) I proposed a theory of schizophrenia in which a pervasive neural integrative defect (schizotaxia) was said to produce, "*on all actually existing social learning regimes*," a personality structure I called, following heterodox analyst Rado, schizotypy. Some critics complained of my using two words when (according to my own theory) they characterized the same people. Here we have a truly undergraduate mistake, failing to distinguish between the *intension* and *extension* of a term. The first is a property, the second is the class of individuals having it. In logician Quine's example, "All animals with a heart have a kidney," does not imply that the words 'heart' and 'kidney' are synonymous! My italics on *actually existing*, suggesting prophylaxis, were simply ignored. I do not suggest that everyone needs philosophy to avoid this elementary blooper; but I *do* maintain that no psychologist who earned a B grade in a logic or philosophy of science course could possibly commit it.

You may be thinking, "But *most* scientists, *most* of the time, must be reasoning rather well, otherwise science shouldn't be as successful as it is." By-passing how successful our science is, especially in the soft fields, that is a *non sequitur*. I am no science basher, and I consider science—if not the only cognitive game in town—surely the best one. But that scientists tend to think better than preachers, politicians, and journalists does not tell us how close they are to reasoning optimally. In educational psychology jargon, I am raising a criterion-referenced rather than a normative question. Of course, ideal cerebration cannot guarantee epistemic success, the whole empirical business being inherently probabilistic. Faust and I conjecture that we function much further below optimum than is generally realized.

While gross fallacies are often motivated, research by clinical, cognitive, and social psychologists shows that a variety of cognitive errors typically occurs in presumably unbiased inferences about affectively neutral matters. The human mind is simply not an efficient computer, except for some special tasks (pattern recognition, language translation) not relevant here. Within the last few years, computer programs have begun to defeat chess masters. When confronted with numerous fallible and conflicting indicators of a theory's verisimilitude, the scientist is forced to employ a subjective regression equation, whether knowingly or not. We can be sure that the scientist's tacit combinatorial function has nonoptimal weights and that they are applied inconsistently (see, e.g., Faust, 1984).

Consider a theory's prediction of numerical observational values. History of science suggests that scientists are favorably impressed when a precise prediction is correct or (despite Popper) *nearly* correct. Sometimes a "near-miss" earns more credit than a "hit" if the "hit" is achieved by making a weak prediction. For example, if your theory of the heritable component of *g* mispredicts a mean foster-child IQ of 117 as 115, most psychologists would give it more credit than my theory gets by correctly predicting an IQ in the interval 105 to 120. In recent papers I have proposed a crude index of confirmation in which the *relative intolerance* of a prediction (as a proportion of the antecedently possible range or *Spielraum*) is multiplied by the *closeness* (Spielraum-relative error) to yield a reciprocally potentiating index of accuracy and risk, to be averaged over experiments, and standardized into the familiar interval [0, 1] by adjusting for best and worst scenarios (see Faust & Meehl, 1992; Meehl, 1990a, 1992a, [2002, 2004]). We can construct other indexes of qualitative diversity, formal simplicity, novel fact predictivity, deductive rigor, and so on. Multiple indexes of theoretical merit could then be plotted over time, intercorrelated, and related to the long-term fate of theories (i.e., in the textbooks as "true," versus discarded). If logicians worry about circularity, we explain to them how psychometric instruments such as the MMPI are derived, bootstrapping item analysis from fallible criteria such as psychiatric diagnoses or interview trait ratings. Fictionists and instrumentalists have no problem here, but scientific realists do, being concerned with verisimilitude. We should not be perfectionist about candidate indexes; their merits and defects will be judged by the *orderliness revealed*, a meta-metacriterion which commands universal assent.

I believe psychologists are more likely to take these steps in cliometric metatheory (Meehl, 1992a) than either historians or philosophers, because of our psychometric skills and our faith in Thorndike's dictum that "if a thing exists, it exists in some amount; and if it exists in some amount, it can be measured." I would urge young cognitive, psychometric, and social psychologists to consider adopting cliometric metatheory as a research career. ('Cliometrics' derives from Clio, the muse of history.) You could be founders of a great enterprise.

I have argued that metatheory can be a help in preventing scientific mistakes, in defending a theory against unsound philosophical criticisms by psychologists using philosophy that is a half-century out of date, and in indexing aspects of theory performance. These roles for metatheory are prophylactic, defensive, and objectifying. Can it be helpful in theoretical creation? Hardly ever, I think, although a few philosophers are taking tentative primitive steps in that direction. But in my own case, the philosophers have served me well even in this. Thirty years ago I was engaged in schizophrenia research, conducting a dozen studies only two of which were published. Why not the others? Because, reading Popper, I realized that these studies, while yielding statistical trends in the expected direction, did not provide a strong, risky test of my theory (or, for that matter, anybody else's). I decided to quit researching schizophrenia until two conditions were met: availability of highly valid indicators, closer in the causal chain to the DNA, and (even good indicators being fallible) availability of a powerful enough taxonomic mathematics (which the cluster algorithms were not). Because theoretical concepts like schizotaxia are defined implicitly via the conjectured nomological network, when that network is incomplete and even the known strands (linking the nodes, which are the theoretical entities) are probabilistic, we deal with Arthur Pap's *open concepts*. (We do not fool ourselves about a so-called "operational definition" of schizotypy, which would be a fake; nor do we say the whole thing is "merely conventional.") How does one subject open concepts to strong Popperian tests? The answer is twofold: (1) Statisticize the openness, (2) Derive theorems relating inferred latent values that serve as numerical consistency tests. Consistency tests are crucial because one lacks an external Gold Standard criterion, the whole enterprise being what Cronbach and I (in 1955) called "bootstrapping" construct validity. I believe, on the basis of extensive Monte Carlo runs and some 13 empirical applications, that I have essentially *solved* the taxometric problem (Meehl, 1973b, Chapter 12, 1992b, p. 131ff; Meehl & Golden, 1982, [1995]). If there is a true latent taxon and there are a few valid indicators, my method will detect taxonicity, estimate the taxon base rate accurately, select the good indicators, optimize the cut on each, estimate the valid and false positive and negative rates, and classify the individuals as accurately as the validities permit. The taxometric problem has existed (in the life sciences) for almost a century. How was I able to solve it? Knowing more mathematics than most clinicians was necessary, as was my extensive clinical experience (lacking in most psychometricians). But what I have learned from the philosophers played a crucial role in thinking about this old, hard problem. If one is clear in the head about both open concepts and Popperian risk, one easily arrives at the solution.

I have one more theme to consider—perhaps more economic and ethical than scientific—the relevance of metatheory for clinical practitioners. One might suppose it to be slight, surely less than for the researcher (clinical or otherwise), but that is incorrect. As I am about to make some challenging remarks, an autobiographical preface is relevant here. Many clinicians perceive me as not a *real* clinician, as one who does not see patients, not on the firing line helping troubled persons, one of those Minnesota types who theorizes about genetics and mathematical models. (I suspect this misperception is held partly because it is comforting for some clinicians thereby to brush aside my arguments and evidence.) The fact is that I have engaged in psychotherapy for over a half-century (I treated my first patient in 1942—I "fixed" him, too!) and have nearly 10,000 hours of therapy practice and, despite my retirement, I currently see several patients. Although my orientation has moved over the years from psychoanalysis to RET, I still have a couch in my office and a picture of Freud on the wall. Do not write me off as an academic who can enjoy the luxury of thinking scientifically because I have no responsibility to care for flesh-and-blood patients.

In relating psychological science to clinical practice, one must first be clear about the demands of the pragmatic context. Whether I am assessing a depressed patient's suicide risk, advising a parole board about recidivism, diagnosing Hoch-Polatin syndrome in a team meeting, or interpreting an analysand's dream, the pragmatic context requires that I *decide what to say*, often without a well-corroborated theory or reliable particulars. Saying nothing, in any of these contexts, is a form of action, as the decision theorists remind us. Nor can I say to the patient "You just lie there on that comfy couch until I get back from checking something in the library," let alone "Come back next month, after I've done my pilot study on the urethral complex." The psychoclinician acts on Jamesian "over-beliefs," as does the pediatrician, lawyer, engineer, or stockbroker. We all have overbeliefs, we could not live without them. You could not choose a graduate school, buy a car, get married (or divorced), let alone invest money or vote for president, if you required all your choices to be derived rigorously from scientific proof reaching a high *p*-value. When there *is* research evidence that a certain MMPI profile pattern is associated 85% of the time with schizophrenia, I rely on that with a clear scientific conscience. But much of the time I lack such quantitative evidence and must utilize whatever plausible mix of clinical experience (my own and that of others I trust), theory, common sense, and intuition I can muster. I make no apologies to my experimental and psychometric colleagues for doing that (Meehl, 1973, 1987). For a more skeptical, scientifically conservative view of clinical experience unsupported by research, see McFall (1991) and Sechrest (1992). As to courtroom testimony by clinicians, it is arguable that high standards of scientific proof are ethically required, since a verdict depriving innocent persons of liberty or property or acquitting rapists, robbers, and murderers is here enforced by the awesome power of the state with its effective monopoly of violence. Discussing that difficult and important question is beyond the scope of this article, but see Ziskin and Faust (1988 [6th ed. by Faust, 2012]), Faust, Ziskin, and Hiers (1991), Underwager and Wakefield (1990), Wakefield and Underwager (1988), and references cited by them.

However, we must distinguish this legitimate invoking of the pragmatic context from a similar-looking but different one that is an abuse. Lacking affirmative research evidence for the validity of a diagnostic or therapeutic procedure is very different from having evidence of its *in*validity. Suppose there is a sizable body of research showing that the Minnesota Tennis-Ball-in-a-Bushel-Basket Projective Technique has negligible validity. Suppose that I can discern no methodological flaws in the studies, or that a meta-analysis yields negligible correlation between methodological weakness and effect size. If I persist in using this test, I am not only an irrational scientist and sloppy scholar, I am also an *unethical practitioner*. It will not do to say "I don't care what the research shows, I am a clinician, so I rely on my clinical experience." Clinical experience may be invoked when it is all we have, when scientific evidence is insufficient (in quantity or quality) to tell us the answer. It is *not* a valid rebuttal when the research answer is negative. One who considers "My experience shows" a valid reply to research studies is self-deceived and must never have read the history of medicine, not to mention the psychology of superstitions. A glaring example of "double standard of methodological morals" is practitioners rejecting the evidence that actuarial prediction equals or excels informal clinical judgment by saying, "But none of these comparative studies deals with exactly the kind of judgments I make," without offering either theoretical argument or empirical data to show that *this* clinician making judgments about *this* predictive domain is somehow different from hundreds of other clinicians operating in a wide variety of judgment domains, many similar to the one in question. If psychology generally proceeded on that methodological basis, no pragmatic or theoretical generalization could ever be corroborated and applied. It is absurd, as well as arrogant, to pretend that acquiring a PhD somehow immunizes me from the errors of sampling, perception, recording, retention, retrieval, and inference to which the human mind is subject. In earlier times, all introductory psychology courses devoted a lecture or two to the classic studies in the psychology of testimony, and one mark of a psychologist was hard-nosed skepticism about folk beliefs. It seems that quite a few clinical psychologists never got exposed to this basic feature of scientific thinking. My teachers at Minnesota (psychologists Hathaway, Paterson, Heron, Skinner, statistician Treloar, philosophers Castell and Feigl) differed widely as to both method and substance. But they shared what Bertrand Russell called the dominant passion of the true scientist—the passion not to be fooled and not to fool anybody else. Only Feigl was a positivist, but all of them persistently asked the two searching questions of positivism: "What do you mean? How do you know?" If we clinicians lose that passion and forget those questions, we are little more than be-doctored, well-paid soothsayers. I see disturbing signs that this is happening and I predict that, if we do not clean up our clinical act and provide our students with role models of scientific thinking, outsiders will do it for us. In this age of litigation, the abuses of "expert" testimony will lead clever, sophisticated lawyers and tough-minded judges to puncture the over-inflated clinical balloon. I have read trial transcripts showing that this is already happening.

Summarizing,

Scientists routinely employ metatheoretic principles, whether calling them "philosophical" or not, in research strategy, interpreting data, and appraising theories.

Explicit discussion of these principles typically occurs in times of intense controversy, theoretical crisis, scientific revolution, and entry into a new domain.

The technical concepts and arguments of philosophers, including their disagreements, are often helpful in such circumstances.

Whether knowledge of formal metatheory helps us in doing "normal science" is not known, but that is researchable. I am sure it has helped me.

Much scientific thinking is of poor quality, and I conjecture it could be improved by explicit metatheoretical education.

Current clinical practice should be critically examined as to scientific quality, and training programs should emphasize rational skepticism, respect for evidence, objectivity, and quantitative thinking more than most of them do.

Conceiving metatheory as the empirical theory of scientific theorizing, the evaluation of competing metatheoretical principles should be conducted by actuarial and psychometric methods applied to random samples of episodes in the history of science. We can help the philosophers at least as much as they can help us.

BIBLIOGRAPHIC NOTE

The literature on metatheory is vast, often difficult and technical, and highly variable in usefulness to a life-scientist. Oldroyd (1986) is a fair-minded over-all survey (though some might want to skip or come back to the long section on the ancients). There is also an excellent introduction by Salmon, Earman, Glymour, Lennox, Machamer, McGuire, Norton, Salmon, and Schaffner (1992) which contains some more technical sections (but doctoral advisors often require able students to learn calculus, matrix algebra, computer programming, or genetics—all of which are "harder" than elementary symbolic logic or traditional epistemology). Sir Karl Popper's emphasis on severe tests and scientific realism has been invaluable to me (Popper, 1959, 1962, 1983; Schilpp, 1974). A clear, deep, forceful defense of Popper (slightly amended) is by Watkins (1984). Important series, but presupposing some knowledge, are *Minnesota Studies in the Philosophy of Science* (1956–) and *Boston Studies in the Philosophy of Science* (1961–). On the knotty problem of appraising theories in psychology, I recommend my own papers (Meehl 1967, 1978, 1990a, 1990b, 1990d). I do not recommend controversial treatments because they presuppose previous knowledge and are frequently careless or tendentious. For psychologists lacking intrinsic philosophical interest but seeking help in their scientific theorizing, the "received view"—a tolerant, amended form of logical empiricism—is the best prescription, and an excellent treatment of it is by Suppe (1977, pp. 7-118). It is still 90% sound, and the 10% error will do little or no harm to the working psychologist. I also recommend Carnap (1966), Hempel (1952), Nagel (1961), and Pap (1962), and there are excellent collections of readings by Boyd, Gasper, and Trout (1991), Brody and Grandy (1989), and Feigl and Brodbeck (1953).

REFERENCES

Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Holt.

Ayer, A.J. (1936). *Language, truth and logic*. New York: Oxford University Press.

Ayer, A.J. (Ed.). (1959). *Logical positivism*. Glencoe, IL: Free Press.

Blumberg, A.E., & Feigl, H. (1931). Logical positivism. *Journal of Philosophy*, *28*, 281-296.

Boyd, R.N., Gasper, P., & Trout, J.D. (Eds.). (1991). *The philosophy of science*. Cambridge, MA: MIT Press.

Broad, C.D. (1933). The "nature" of a continuant. In his, *Examination of McTaggart's Philosophy*, Vol. I (pp. 264-278). Cambridge, UK: Cambridge University Press.

Brody, B., & Grandy, R. (Eds.). (1989). *Readings in the philosophy of science* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Carnap, R. (1936–1937). Testability and meaning. *Philosophy of Science*, *3*, 420-471; *4*, 2-40. Reprinted with corrigenda and additional bibliography, New Haven, CT: Yale University Graduate Philosophy Club, 1950.

Carnap, R. (1945). The two concepts of probability. *Philosophy and Phenomenological Research*, *5*, 513-532.

Carnap, R. (1966). *Philosophical foundations of physics*. New York: Basic Books.

Cohen, R.S. (Ed.). (1981). *Herbert Feigl. Inquiries and provocations: Selected writings 1929-1974*. Boston, MA: D. Reidel.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

Efron, D., & Foley, J.P., Jr. (1937). A comparative investigation of gestural behavior patterns in Italian and Jewish groups living under different as well as similar environmental conditions. *Zeitschrift Sozialforsch*, *6*, 151-159. Reprinted in T.M. Newcomb & E.L. Hartley (Eds.), *Readings in social psychology* (pp. 33-40). New York: Holt, 1947.

Faust, D. (1984). *The limits of scientific reasoning*. Minneapolis, MN: University of Minnesota Press.

[Faust, D. (2012). *Coping with psychiatric and psychological testimony* (6th ed.). NY: Oxford University Press.]

Faust, D., & Meehl, P.E. (1992). Using scientific methods to resolve enduring questions within the history and philosophy of science: Some illustrations. *Behavior Therapy*, *23*, 195-211.

Faust, D., Ziskin, J., & Hiers, J.B., Jr. (1991). *Brain damage claims: Coping with neuropsychological evidence*. (Vols. 1–2) Los Angeles, CA: Law & Psychology Press.

Feigl, H. (1969). The *Wiener Kreis* in America. In D. Fleming & B. Bailyn (Eds.), *The intellectual migration 1930-1960* (pp. 630-673). Cambridge, MA: Harvard University Press. Reprinted in R.S. Cohen (Ed.), *Herbert Feigl. Inquiries and provocations: Selected writings 1929-1974* (pp. 57-94). Boston, MA: D. Reidel, 1981.

Feigl, H., & Brodbeck, M. (Eds.) (1953). *Readings in the philosophy of science*. New York: Appleton-Century-Crofts.

Freud, S. (1920/1955). *Beyond the pleasure principle*. In J. Strachey (Ed. & Trans.), *Standard edition of the complete psychological works of Sigmund Freud* (Vol. 18, pp. 7-64). London: Hogarth. (Original work published 1920)

Gilman, D. (1992). What's a theory to do… with seeing? Or, some empirical considerations for observation and theory. *British Journal for the Philosophy of Science*, *43*, 287-309.

Glymour, C. (1980). *Theory and evidence*. Princeton, NJ: Princeton University Press.

Griffith, A.V., & Fowler, D. (1960). Psychasthenic and hypomanic scales of the MMPI and reaction to authority. *Journal of Counseling Psychology*, *7*, 146-147.

Grove, W.M. (1986, May). *Clinical inference from psychological tests: Last nails in the coffin*. Paper presented at the meeting of the Minnesota Psychological Association, Minneapolis, MN.

[Grove, W.M., Zald, D.H., Lebow, B.S., Snits, B.E., & Nelson, C.E. (2000). Clinical vs. mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19-30.]

Hempel, C.G. (1952). Fundamentals of concept formation in empirical science. In *International Encyclopedia of Unified Science*. Vol. II, No. 7. Chicago, IL: University of Chicago Press.

King, P., & Steiner, R. (Eds.) (1991). *The Freud-Klein controversies*. New York: Tavistock/Routledge.

Kordig, C.R. (1971). The comparability of scientific theories. *Philosophy of Science*, *38*, 467-485.

Kuhn., T.S. (1970). *The structure of scientific revolutions* (2nd ed). *International Encyclopedia of Unified Science*, *2*(2). Chicago, IL: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-195). Cambridge, UK: Cambridge University Press. Reprinted in J. Worrall & G. Currie (Eds.), *Imre Lakatos: philosophical papers. Vol. I. The methodology of scientific research programmes* (pp. 8-101). New York: Cambridge University Press, 1978.

Laudan, L., Donovan, A., Laudan, R., Barker, P., Brown, H., Leplin, J., Thagard, P., & Wykstra, S. (1986). Scientific change: P hilosophical models and historical research. *Synthese*, *69*, 141-223.

Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy*, *67*, 427-446.

Mackie, J.L. (1965). Causes and conditions. *American Philosophical Quarterly*, *2*, 1-20.

Mahl, G.F. (1987). *Explorations in nonverbal and vocal behavior*. Hillsdale, NJ: Erlbaum.

Mahoney, M.J., & Kimper, T.P. (1976). From ethics to logic: A survey of scientists. In M.J. Mahoney, *Scientist as subject: The psychological imperative* (pp. 187-193). Cambridge, MA: Bollinger.

Maxwell, G. (1962). The ontological status of theoretical entities. In H. Feigl & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science: Vol. 3. Scientific explanations, space, and time* (pp. 3-27). Minneapolis, MN: University of Minnesota Press.

Maxwell, G. (1970). Structural realism and the meaning of theoretical terms. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science: Vol. 4. Analyses of theories and methods of physics and psychology* (pp. 181-192). Minneapolis, MN: University of Minnesota Press.

McFall, R.M. (1991). Manifesto for a science of clinical psychology. *The Clinical Psychologist*, *44*, 75-88.

Meehl, P.E. (1954/1996). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press. Reprinted with new Preface, 1996, by Jason Aronson, Northvale, NJ. Available at http://www.tc.umn.edu/~pemeehl/

Meehl, P.E. (1962) Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, *17*, 827-838. Available at http://www.tc.umn.edu/~pemeehl/

Meehl, P.E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*, 103-115. Reprinted in D. E. Morrison & R. E. Henkel (Eds.), *The significance test controversy* (pp. 252-266). Chicago, IL: Aldine, 1970. Available at http://www.tc.umn.edu/~pemeehl/

Meehl, P.E. (1972). Specific genetic etiology, psychodynamics, and therapeutic nihilism. *International Journal of Mental Health*, *1*, 10-27.

Meehl, P.E. (1973). *Psychodiagnosis: Selected papers*. Minneapolis, MN: University of Minnesota Press.

Meehl, P.E. (1977). Specific etiology and other forms of strong influence: Some quantitative meanings. *Journal of Medicine and Philosophy*, *2*, 33-53.

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. Available at http://www.tc.umn.edu/~pemeehl

Meehl, P.E. (1983). Subjectivity in psychoanalytic inference: The nagging persistence of Wilhelm Fliess's Achensee question. In J. Earman (Ed.), *Minnesota studies in the philosophy of science: 10. Testing scientific theories* (pp. 349-411). Minneapolis, MN: University of Minnesota Press. Reprinted in *P.E. Meehl, Selected philosophical and methodological papers* (pp. 284-337). (C.A. Anderson & K. Gunderson, Eds.) Minneapolis, MN: University of Minnesota Press, 1991.

Meehl, P.E. (1986). Trait language and behaviorese. In T. Thompson & M.D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 315-334). Hillsdale, NJ: Erlbaum.

Meehl, P.E. (1987). Theory and practice: reflections of an academic clinician. In E.F. Bourg, R.J. Bent, J.E. Callan, N.F. Jones, J. McHolland, & G. Stricker (Eds.), *Standards and evaluation in the education and training of professional psychologists* (pp. 7-23). Norman, OK: Transcript Press. Available at http://www.tc.umn.edu/~pemeehl/

Meehl, P.E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant using it. *Psychological Inquiry*, *1*, 108-141, 173-180. Available at http://www.tc.umn.edu/~pemeehl/

Meehl, P.E. (1990b). *Corroboration and verisimilitude: Against Lakatos' "sheer leap of faith"* (Working Paper, MCPS–90–01). Minneapolis, MN: University of Minnesota, Center for Philosophy of Science. Available at http://www.tc.umn.edu/~pemeehl/

Meehl, P.E. (1990c). Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *Journal of Personality Disorders*, *4*, 1-99. Available at http://www.tc.umn.edu /~pemeehl/

Meehl, P.E. (1990d). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195-244. Also in R.E. Snow & D. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 13-59). Hillsdale, NJ: Erlbaum, 1991. Available at http://www.tc.umn.edu/~pemeehl/

Meehl, P.E. (1992a). Cliometric metatheory: The actuarial approach to empirical, history-based philosophy of science. *Psychological Reports*, *71*, 339-467. [This is the first of three publications that Meehl originally intended to publish as a book. It was followed by: "Cliometric metatheory II: Criteria scientists use in theory appraisal and why it is rational to do so." *Psychological Reports*, 2002, *91*, 339-404; and, published posthumously, "Cliometric metatheory III: Peircean consensus, verisimilitude, and asymptotic method." *British Journal for the Philosophy of Science*, 2004, *55*, 615-643. —LJY] All three available at http://www.tc.umn.edu/~pemeehl/

Meehl, P.E. (1992b). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*, *60*, 117-174. Available at http://www.tc.umn.edu /~pemeehl/

[Meehl, P.E. (1995) Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*, 266-275. Available at http://www.tc.umn .edu/~pemeehl/]

Meehl, P.E. (in pareparation). When one's metatheory matters, it can matter a lot: The case of path analysis. [Published as: Meehl, P.E., & Waller, N.G. (2002). The path analysis controversy: A new statistical approach to strong appraisal of verisimilitude. *Psychological Methods, 7*, 283-300; Waller, N.G., & Meehl, P.E. (2002). Risky tests, verisimilitude, and path analysis. *Psychological Methods, 7*, 323-337.]

Meehl, P.E., & Golden, R. (1982). Taxometric methods. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127-181). New York: Wiley.

Morrison, D.E., Henkel, R.E. (Eds.) (1970). *The significance test controversy*. Chicago, IL: Aldine.

Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace & World.

Oldroyd, D. (1986). *The arch of knowledge: An introductory study of the history of the philosophy and methodology of science*. New York: Methuen.
Palermo, D.S., & Jenkins, J.J. (1964). *Word association norms: Fourth grade through college*. Minneapolis, MN: University of Minnesota Press.
Pap, A. (1953). Reduction-sentences and open concepts. *Methodos*, *5*, 3-30.
Pap, A. (1958). *Semantics and necessary truth*. New Haven, CT: Yale University Press.
Pap, A. (1962). *An introduction to the philosophy of science*. New York: Free Press.
Passmore, J. (1967). Logical positivism. In P. Edwards (Ed.), *Encyclopedia of philosophy* (Vol. 5, pp. 52-57). New York: Macmillan.
Popper, K.R. (1959). *The logic of scientific discovery*. New York: Basic Books. (Original work published 1934)
Popper, K.R. (1962). *Conjectures and refutations*. New York: Basic Books.
Popper, K.R. (1967). A revised definition of natural necessity. *British Journal for the Philosophy of Science*, *18*, 316-321.
Popper, K.R. (1983). *Postscript to the logic of scientific discovery. Vol. I: Realism and the aim of science*. Totowa, NJ: Rowman and Littlefield.
Reichenbach, H. (1938). *Experience and prediction*. Chicago, IL: University of Chicago Press.
Salmon, M.H., Earman, J., Glymour, C., Lennox, J.G., Machamer, P., McGuire, J.E., Norton, J.D., Salmon, W.C., & Schaffner, K.F. (1992). *Introduction to the philosophy of science*. Englewood Cliffs, NJ: Prentice Hall.
Schilpp, P. (Ed.), (1974). *The philosophy of Karl Popper*. LaSalle, IL: Open Court.
Sechrest, L. (1992). The past future of clinical psychology: A reflection on Woodworth (1937). *Journal of Consulting and Clinical Psychology*, *60*, 18-23.
Skinner, B.F. (1948). *Walden Two*. New York: Macmillan.
Suppe, F. (Ed.). (1977). *The structure of scientific theories* (2nd ed.). Urbana, IL: University of Illinois Press.
Underwager, R., & Wakefield, H. (1990). *The real world of child interrogations*. Springfield, IL: Thomas.
Wakefield, H., & Underwager, R. (1988). *Accusations of child sexual abuse*. Springfield, IL: Thomas.
Walker, E., & Lewine, R.J. (1990). Prediction of adult-onset schizophrenia from childhood home movies of the patients. *American Journal of Psychiatry*, *147*, 1052-1056.
Watkins, J.W.N. (1984). *Science and scepticism*. Princeton, NJ: Princeton University Press.
Weinberg, J.R. (1936). *An examination of Logical Positivism*. New York: Harcourt, Brace.
Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*, 557-585.
Ziskin, J., & Faust, D. (1988). *Coping with psychiatric and psychological testimony* (4th ed.) Vols. 1-3. Marina del Rey, CA: Law and Psychology Press.